PB84-222363

National Statistical Assessment of Rural Water Conditions. Volume 4

Cornell Univ., Ithaca, NY

Prepared for

Environmental Protection Agency, Washington, DC

Jun 84

U.S. Department of Commerce
National Technical Information Service
NTIS

United States
Environmental Protection
Agency

Office of
Drinking Water (WH-550)
Washington, DC 20460

EPA 570/9-84-004(d)
June 1984

PB84-222363

Water

EPA

National Statistical Assessment of Rural Water Conditions

Volume IV

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.
EPA 570/9-84-004-D

2.

3. RECIPIENT'S ACCESSION NO.
PB8 4 222363

4. TITLE AND SUBTITLE
National Statistical Assessment of Rural
Water Conditions, Volume -IV

5. REPORT DATE
June 1984

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Department of Rural Sociology
Cornell University

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Cornell University
Department of Rural Socialocy

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS
U.S. Environmental Protection Agency
Office of Drinking Water (WH-550)
Washington, DC 20460

13. TYPE OF REPORT AND PERIOD COVERED

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES
Four Volume Set

16. ABSTRACT

This study considered five dominant dimensions of the status of domestic water: quality, quantity, availability, cost and affordability. Rural residents were asked about health effects but the results were modest in that very few rural residents reported adverse health conditions which they associated with the water supply.

With enactment of the Safe Drinking Water Act of 1974, Congress set in motion two major efforts to develop systematic, current data on rural water supplies across the nation. First, in response to growing concern with the quality of drinking water and its effects on human health, the Safe Drinking Water Act provided for a uniform, national set of water quality standards and extended the monitoring and regulatory responsibility of the US Government over smaller water supplies. Second, the Act mandated a one-time national statistical assessment of the current status of rural domestic water characteristics. This document fulfills that mandate.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Rural water drinking water Rural water conditions water supply systems water quality water quantity		

18. DISTRIBUTION STATEMENT
Open distribution

19. SECURITY CLASS *(This Report)*
non-sensitive

20. SECURITY CLASS *(This page)*

21. NO. OF PAGES
316

22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

IX

Introduction to NSA Analysis Chapters

The previous chapters and sections of this report primarily have been concerned with the scope, dimensions, and types of considerations addressed in the National Statistical Assessment of Rural Water Conditions, as well as presenting a detailed characterization of the status of rural domestic water and the people who consume it. Chapters III through VIII have described the findings for hundreds of variables (observations and measurements) taken on rural households or water systems during the summer and fall of 1978. The data were first compiled for the nation as a whole. Then the same data were presented for the four broad Census regions of the country, for households located inside and outside SMSA counties, by size of place, and size of the water supply system from which the household obtained its water.

Those chapters were conceived and prepared to describe differences between households or systems as manifested by individual variables and no attempt was made to relate any variable to another to determine whether they covaried systematically. Hence the preceding material constituted the descriptive, nonanalytical portion of the report.

In the chapters that follow, Chapters X through XIII, the focus and concern shifted away from the extensive description of univariate data to exploration of various bivariate and multivariate relationships. The emphasis of the following chapters, then, was analytical in nature since mainly relationships between variables or between composites and variables were considered.

With emphasis on the types and strength of relationships among variables or composites, clearly less detail about the individual variables involved was presented, this having been accomplished earlier. Accordingly, many of the variables which had been described independently of one another previously were interwoven into composite summary measures or scores. Many of these composites were described in Chapter VI; others were introduced for the first time in the text developed in the next set of chapters.

Each of the next chapters restricted attention to only one major dimension of the status of rural drinking water. Chapter X was restricted to an analysis of water quality, Chapter XI was limited to an analysis of the quantity of water available and consumed at the rural household, Chapter XII explored the availability dimension and Chapter XIII evaluated the cost of water to the households. An advantage of this arrangement, which restricted the investigation in each chapter to a unique status component, was that a more complete analysis of that dimension could be accomplished and communicated.

In each of these four chapters, however, more than one measure of the dimension was analyzed. To conduct a comprehensive analysis of water quality, eight variables or composites were employed. Selected as dependent variables there, and separately analyzed, were total coliform, turbidity, lead, nitrates, as well as a health-related composite, an aesthetic/economic composite, a general water quality composite, and a perceived water quality composite. Likewise, a comprehensive evaluation of factors relating to water quantity involved employing two dependent variables—billed quantity and the residents' perceptions regarding the quantity of water supplied to the household. Similarly, the assessment of relationships involving the rural household's availability to a water supply resulted in the choice of two measures as dependent variables—an accessibility composite and an index of the supply system's reliability. Finally, in Chapter XIII, two measures were employed—the monthly cost of the household supply and the per thousand gallon cost to the user. Extensive descriptions of all of these dependent variables and composites were given previously in Chapters V and VI.

Since interest among professionals working in water supply has been centered on knowledge of relationships involving each of these variables or composites separately, and because the relationships among these dependent variables were weak, these conditions were judged ample evidence and reason to conceive and treat each dependent variable as a distinct aspect of the status of rural water conditions. This meant that each dependent variable should be modeled separately rather than via a multivariate technique like canonical regression which assumes high correlation among the dependent variables. Hence models of relationships between each dependent variable and sets of relevant household characteristics, supply attributes, and supply system variables were developed and presented independently. That is, relationships between coliform measures and household, supply, and system variables were modeled independently from turbidity, and so forth.

GOALS OF ANALYSIS

The goals of the analyses presented in the following chapters were more limited than some other studies as large, comprehensive, and as detailed as the NSA. No attempt was made to claim that these data and sets of relationships explain variations in the status of the quality, quantity, availability, and cost of drinking water in the rural US. More specifically, these analyses do not assert that all relevant variables have been examined nor does it assert cause and effect relationships. To even partially establish causality would require the inclusion of other relevant variables into the analysis like: geologic and hydrologic information, information from a program of quality and quantity monitoring of aquifers and surface sources, more detailed data on the construction and alterations of the supply, age and condition of the household plumbing, multiple sampling of the household's water, knowledge of pricing schedules used by supply systems, plus many other variables which the NSA did not include. Also required to assert causality from a survey study would be longitudinal measures on the dependent variables to assess whether changes in the dependent variable distributions could be plausibly ascribed to the independent variables.

While a simple correlation between two variables measures the degree to which they covary, it does not necessarily signify that one variable determines or produces variations in the other. A major implication of this presumption is that, unlike experimental work, the consequences of statistically "manipulating" one variable does not directly produce a change in another even though the two variables may be systematically interrelated. For example, showing the association between a predictor (independent) variable and a particular condition on the predicted variable, such as excessive coliform organisms, does not assure that the latter condition can be altered by changing the level of the other "independent" variable with which it is correlated.

On the other hand, when an independent variable, like whether the water supply system chlorinates, demonstrates a substantial degree of association with the dependent variable being analyzed (i.e., the number of coliform colonies in the water), this information should prove useful to researchers and practitioners alike, because the independent variables being used in the analytical model tend to be more readily observed than the information contained in the dependent variables —quality, quantity consumed, system reliability, etc. Hence, knowledge of these more readily observed independent variables and knowledge of their relationship with the dependent variable makes it possible to improve predictions about the status of a household's drinking water than without this knowledge.

Hence, the main goal of these analyses was limited to establishing regularity of covariation—i.e., the identification of variables that demonstrated a systematic relationship (correlation or regression) with the dependent variables. Those variables which had a sufficient degree of correlation to account for some of the variability or variance in the dependent variables were retained for consideration, the nature of their relationships modeled, and interpretations made. Mention was also made of those variables which did not demonstrate a sufficient relationship to warrant further analysis even though they were initially included because of their indicated importance in (1) previous research, (2) conventional wisdom of professionals working in the water supply area, or (3) import for regulatory and financial programs.

A second goal of analysis was exploratory. Because the NSA was the first systematic national assessment of rural domestic water conditions, the analysis goals were more exploratory than confirmatory in nature; no attempt was made to test traditional hypotheses. This latter approach would have required more certainty of knowledge from previous research that all the relevant factors had been included in the models (and no irrelevant ones). However, since this certainty of knowledge does not exist for many of the dimensions of the status of drinking water examined in the NSA, attempts at classical hypothesis testing and definitive explanations seemed premature.

Consistent with these two narrowly circumscribed objectives of the NSA, each analysis presented in the following chapters was structured to investigate patterns of covarying relationships (correlations and regressions) that may be instrumental in identifying households with a tendency to have problematic water supplies. The information produced by each analytical segment may be useful in delineating these households, but it does not provide a sufficient basis for devising and implementing strategies for mitigating the causes of the problems. Once households with the stipulated condition are located, further investigation may be required to identify the source of the condition and to determine how it can be eliminated.

ANALYSIS FRAMEWORK—MULTIPLE REGRESSION

Several constraints and desired objectives for the analysis dictated the selection of multiple regression as the common analytical framework employed in the ensuing chapters. Some of these constraints were: (1) wide range of content of the dependent variables—ranging from water quality measures to cost, (2) complex phenomena encapsulated in the dependent variables—even a concept like availability was more complicated than originally envisioned, (3) irregularity of the distributions of some of the dependent variables, (4) use of several independent variables in modeling, (5) use of both categorical and continuous independent variables, (6) unequal number of cases in various categories or score intervals of the independent variables, and (7) need for a standard, common analytical framework to ease communication of results.

First, the range of content encapsulated by the dependent variables was quite heterogeneous. Because the dependent variables are so different in nature, there was no reason to assume that any given independent variable would automatically be systematically related to them all. Secondly, because of the complex phenomena summarized (or incorporated) in many of the composites, it seemed advisable to assume that no one independent variable would demonstrate such a strong relationship to the dependent variable that it alone was sufficient to account for all or most of the variability in the dependent variable. Instead it was assumed that relationships between each dependent variable and a set of independent variables, considered simultaneously, would perform better in terms of the analysis goals.

Third, some of the variables selected as dependent and proper indicators of the dimensions of status of rural drinking water turned out to be essentially a dichotomy (a two-category variable). The perceived quantity composite, for example, was distributed such that approximately 81 percent of all occupied rural households were supplied with sufficient quantity of water throughout the year (see Chapter V). Hence, 81 percent had the same score on that dependent variable, while only 19 percent (the remaining households) had a score that varied from this. Only a slightly improved distribution of households occurred for the reliability composite, where approximately three-fourths of all occupied rural households had a score of zero, to which could be added about 7 percent with a score of one. Hence, roughly 82 percent had essentially the same score on that composite with the remaining 18 percent assuming a range of scores. For analysis purposes, then, each of these two dependent variables appeared to be best conceived as a two valued variable. One of the flexible properties of regression is that it can properly accommodate the analysis of dichotomous dependent variables because it shares the property of least squares optimization of one linear function with discriminant analysis, which would normally be the statistical model selected for analyzing a dependent variable consisting, effectively, of two categories.

Fourth, multiple regression is an entire collection of flexible statistical techniques which allows one to simultaneously study the relationships between a single dependent variable with a host of variables designated as independent. This was an essential property in light of the above stated objectives, but equally important for another reason. Many times the relationship which is observed between two variables becomes refined (altered) when studied in the context of a third (or more) other variables with which they are both associated. More importantly, it is generally the refined relationship, which becomes elaborated when controlling for the common influence of one or more other variables, that is closer to the "true" relationship.

Fifth, multiple regression is equally valid for independent variables which are categorical as well as continuous in distribution. Sixth, it can easily accommodate unequal numbers of cases (households) located in different categories of the independent variable. Since the NSA involved unequal cell sizes and analysis required the exploration of systematic associations between the dependent variables and both categorical (discrete) and continuous independent variables, these properties made regression the preferred analytical strategy. The last factor in the decision on an analytical framework was the benefit of easier communication of results and interpretations when utilizing a standard, common analytical technique. All of these considerations, plus other more technical ones about statistical inference, congruence between data and statistical models, and power/ efficiency, led to the selection of multiple regression.

ASSUMPTIONS OF MULTIPLE REGRESSION AND CONFORMABILITY OF DATA PROPERTIES

Multiple regression shares certain assumptions about the properties of the data with kindred techniques encompassed under what is called the general linear model. Similar to analysis of variance models, covariance models and discriminant models, multiple regression assumes that (1) both the independent and dependent variables have been measured without error, (2) except for discriminant models, the dependent variable is continuous and ordered, (3) the model has been correctly specified—no relevant independent variables have been excluded and no irrelevant independents have been included, (4) the error term is normally distributed and the expected value of the error term is zero, (5) the variance of the error term is constant for all values or combination of values of the independent variables, and (6) the error terms are uncorrelated. To the extent that the data depart from conformability with the assumptions, some degree of bias is introduced in the estimation of the regression coefficients. On the other hand, regression analysis is robust in estimation when only a few of these assumptions have been violated.

Few data sets gathered from field surveys perfectly conform to all the assumptions. Likewise, the NSA data set probably violates certain of these assumptions to some degree. However, by the use of transformation techniques and judicious selection of more specialized techniques of analysis, the NSA data have been properly analyzed with a minimal deviation from assumptions. For example, when various transformations of the perceived quantity and reliability composites failed to provide optimal statistical properties, the decision was made to employ discriminant analysis in examining relationships with those dependent variables, as it would be the more proper model to utilize.

ALTERNATIVE ANALYSIS MODELS

Some controversy exists among statisticians regarding how to proceed when data fail to conform to the assumptions of the regression model. One group of theorists and most practitioners (including applied researchers) have tended to rely upon regression techniques anyway, recognizing that the weight of evidence indicates that parameter estimation and inference remains robust under slight deviations. Indeed, studies on the subject suggest that the most common result of these violations is to produce conservative estimates—i.e., tend to underestimate the importance of the relationship. On the other hand, the statistical purist argues that when the assumptions of regression are not met, the analyst should employ a nonparametric or distribution-free statistical model for studying relationships, albeit less powerful and efficient.

The stance of the purist causes several difficulties for the researcher, however. First, the use of nonparametric techniques means a loss of valuable, hard won information, as these techniques are, at most, sensitive to differences in the order of score values, remaining insensitive to information about magnitude of differences. Second, they are less powerful and efficient, requiring a larger sample size to achieve the same level of statistical significance. Third, and more importantly, they cannot readily handle more than three variable relationships simultaneously without becoming overwhelmingly cumbersome to manipulate and interpret. Some of the distribution-free models can't even summarize three variable relationships. Fourth, other nonparametrics do not have well established theoretical sampling distributions, which are essential for interpretation of the strength of a relationship. Fifth, each of these models carry their own set of assumptions about the nature of the data properties and though these assumptions are less restrictive in some senses the models tend to have been developed to meet a very circumscribed set of conditions which makes them quite restrictive in fact. The result of this last constraint is that the researcher would be forced to use a different model, which produces different summary coefficients, for every particular circumstance imposed by the data properties of the various variables. That condition, in turn, would make communication of interpretations extremely tedious, if not impossible.

Technically then, the trade-off involved in the decision is between using a set of less powerful and flexible nonparametric models, with attending loss of information and precision, or employing a more powerful, information preserving parametric technique like regression in spite of slight deviations from the assumptions. The authors of this report chose to use regression analysis.

What this choice implies for some of the analyses presented is that some of the predictions, some of the coefficients, and some of the systematic variance accounted for could be slightly inaccurate. However, the gains of using a common analysis framework and the comparability of interpretations tendered therefrom more than compensated for these difficulties.

Given the exploratory nature of the whole analysis at this stage and with limited knowledge about rural household water conditions, this was deemed sufficient justification for proceeding with regression. Regression analysis is being utilized as a heuristic device for exploratory modeling of relationships between water quality, quantity, availability, and cost, and various characteristics of the supply, the supply system, and the households involved. Used heuristically, regression techniques will properly guide the selection of variables (from among the set available in the NSA) which demonstrate patterns of systematic relationship with the dependent variables of interest. These techniques will allow the assessment of the relative importance of the retained independent variables, by investigating the strength of their relationship in the context of other independent variables, including the sampling design variables—region, SMSA, and size of place.

STRATEGY FOR MODELING RELATIONSHIPS WITH THE DEPENDENT VARIABLES

For each dependent variable analyzed in the following chapters, the authors set forth a list of variables selected from those contained in the NSA data set which were expected to bear some meaningful relationship with the dependent variable and, therefore, evidence some pattern of systematic variation. The selection of these candidate independent variables was based on several considerations including: (1) previous findings from other studies; (2) experience with the NSA data set provided through the development of the descriptive chapters (Chapters III through VIII); (3) questions posed by the EPA staff; and (4) suggestions from professionals working in the water quality and supply areas. To facilitate the analyses, these variables were then grouped into three distinct sets according to whether they were judged to more dominantly reflect an aspect of the rural household and its occupants, reflect an attribute of the supply, or predominately measure a supply system characteristic. The assumption underlying the formation of each set was that its component variables were more homogeneous in content, kind and magnitude of influence, or relationship with the dependent variable, than variables placed in another set. These three sets (household, supply, and system) were chosen based on the authors' judgment that they represented the primary spheres of influence on rural water quality, quantity, availability, and cost.

Examples of independent variables grouped into the household set are number of occupants of the household, length of residence, income, the head's age, education, ethnicity, and occupants' perceptions of the supply. For a comprehensive list of the household variables included, consult Table IX-1. Examples of independent variables included in the supply set are number of water treatment devices, means of transporting water, and type of supply. For an inclusive list, see Table IX-2. Examples of independent variables included in the supply system set are system age, number of connections, number of operators, ownership of system, plus various economic and operational attributes. Again, for more detail, consult Table IX-3.

The reader will note that within each set, some variables have been distinguished as being continuous in property, whereas others have been designated as categorical. This distinction carries much importance in terms of the way the variable was treated in regression analysis as alluded to earlier and is emphasized here to signal that the results were presented somewhat differently in the ensuing analytical chapters.

Tables IX-1 through IX-3 provide only a general description of the entire set of independent variables whose relationships with the dependent variables will be modeled. The special manner in which the variable was treated within the context of the particular dependent variable being analyzed will be discussed in each of the following chapters.

USE OF CATEGORICAL INDEPENDENT VARIABLES

Though categorical independent variables can be accommodated within the multiple regression analytical framework, they require special treatment. A

Table IX-1
Household Attributes Used in Analytical Chapters

Variable Name	Short Description	Chapter Containing Original Discussion
Continuous Variables		
AGE	Age of the household head	III
EDUCATION	Number of years of schooling completed by the household head	III
INCOME	Total household income	III
PEOPLE	Number of residents in a household	III
RESIDENCE	Length of residence of the household head	III
UNITS	Number of separate housing units in the structure	III
YEARS SERVED	Number of years the household respondent had been served by the water supply system	III
AFFORDABILITY	The ratio of the cost of water to total household income, applicable only to households on community systems	V
DEVICES	Number of water-using fixtures and devices present in a household	III
DISLIKED	Number of disliked supply characteristics	See text, Chapters XI, XII, XIII
LIKED	Number of liked supply characteristics	See text, Chapters XI, XII, XIII
OUTDOOR USES	Number of outdoor uses of water	See text, Chapter XI
PERCEIVED QUALITY	Score on a composite measure of perceived quality related to odor, taste, cloudiness, color, and sediment; Lower scores reflect higher perceived quality	VI
QUANTITY	Number of liters of water used per month	V

Table IX-1 continued

Variable Name	Short Description	Chapter Containing Original Discussion
Continuous Variables (continued)		
PERCEIVED QUANTITY	Score on a composite measure of perceived quantity based on the amount of water supplied and the consistency of the amount supplied, as reported by the household respondent; Lower scores reflect higher perceived quality	V
SATISFACTION	Score on a summated scale of household satisfaction with the water supply in regard to odor, taste, color, cloudiness, sediment, pressure, and quantity; Higher scores reflect greater satisfaction	See text, Chapters XI, XIII
UNIT COST	Cost of 1,000 gallons of water	V
COST PER MONTH	Cost of water for one month	V
Categorical Variables		
ETHNICITY	Ethnicity of household head White Nonwhite (includes Black, Hispanic, American Indian, Alaskan Native, Asian, and Pacific Islander)	III
MOBILE HOME	Type of construction of the housing unit Mobile home Not mobile home	III
TENURE	Type of housing tenure Own Rent Other	III
WORKING	Work status of household head Working full time Working part time (includes students) Keeping house Retired Unemployed	III

Table IX-1 continued

Variable Name	Short Description	Chapter Containing Original Discussion
Categorical Variables (continued)		
PERCEIVED COST	Perceived cost of the household water supply No cost Inexpensive Reasonable Expensive Very expensive	V
PREFERRED SYSTEM	Household preference for present system or different kind of system Present system Different system	III
PREFERRED OWNER	Household preference for system ownership Private groups and individuals (includes individuals, cooperatives, and Indian tribes) Private firm Public Other (includes no owner)	III
PREFERRED SIZE	Household preference for size of system Individual Intermediate Community	III
WILLINGNESS TO PAY	Willingness to pay more, the same, or less for an ideal water supply than the cost of the present supply More Same Less	V

Table IX-2

Supply Attributes Used in Analytical Chapters

Variable Name	Short Description	Chapter Containing Original Discussion
Continuous Variables		
ACCESSIBILITY	Score on an index of accessibility based on information about recorded pressure and the distance between the extraction point and the household; Lower scores indicate a more accessible supply	V
AESTHETIC/ ECONOMIC QUALITY	Score on a specialized index of water quality based on laboratory measurements of constituents having aesthetic or economic effects (the five with secondary MCLs); Lower scores indicate higher quality	VI
COLIFORM	Laboratory count of total coliforms per 100 milligrams of household water	V
GENERAL QUALITY	Score on a generalized index of water quality based on laboratory measurements of 23 constituents; Lower scores indicate higher quality	VI
HEALTH QUALITY	Score on a specialized index of water quality based on laboratory measurements of the eleven constituents studied which had primary MCLs; Lower scores indicate higher quality	VI
LEAD	Milligrams per liter of lead in household water	V
NITRATE-N	Milligrams per liter of Nitrate-N in household water	V
RELIABILITY	Score on an index reflecting the number of minor and severe water supply breakdowns during the year preceding the NSA; Lower scores reflect higher reliability	V
TOTAL TREATMENT	Number of treatment devices present at the household together with the number present at the system	See text, Chapters X, XII, XIII
TURBIDITY	NTU count of household water	V

Table IX-2 continued

Variable Name	Short Description	Chapter Containing Original Discussion
Continuous Variables continued		
WELL AGE	Number of years since construction or last major alteration of the well	VII
WELL DEPTH	Depth of the well	VII
Categorical Variables		
CONVEYANCE	Means used for transporting water to the household Pressurized public supply Electric pump Other (includes hand pump, diesel- or gasoline-driven pump, artesian pressure, gravity flow, hand-carrying, and hauling)	VII
SUPPLY TYPE	Type of water supply serving the household Community supply Well Other (includes spring, surface water supply, cistern, hauled, and purchased bottled)	IV
WELL TYPE	Type of well construction Drilled Driven Jetted Bored Dug	VII

Table IX-3

System Attributes Used in Analytical Chapters

Variable Name	Short Description	Chapter Containing Original Discussion
Continuous Variables		
CONNECTIONS	Number of service connections	VII
DEBT RETIREMENT	Portion of revenue allocated to debt retirement	VIII
LINES	Number of miles of distribution lines in the system	VII
O & M	Portion of revenue allocated to operation and maintenance costs	VIII
OPERATORS	Number of people who operate the system, excluding office and clerical staff	VII
PERFORMANCE	Operating performance of system	VII
REVENUE	Annual revenue of system	VIII
REVENUE/CXN	Annual revenue per connection	See text, Chapters XI, XII, XIII
SYSTEM AGE	Number of years since construction of system or since last major alteration of system	VII
SERVICE YEARS	YEARS SERVED for single connection systems; SYSTEM AGE for multiple connection systems	See text, Chapter X
Categorical Variables		
OWNERSHIP	Type of system ownership Private groups and individuals (includes individuals, cooperatives, and Indian tribes) Private firm Public Other (includes no owner)	VII

Table IX-3 continued

Variable Name	Short Description	Chapter Containing Original Discussion
Categorical Variables (continued)		
SOURCE	Source of water Groundwater Surface water Purchased Mixed (multiple sources)	IV, VII
WATER TEST HISTORY	Composite history of testing of the water supply, combining testing done by the household and testing done by the system No testing Testing by the household Testing by the system, or by both system and household	See text, Chapter X

categorical variable is one in which cases (households) differ in type or kind. Each household is assigned to one of a set of mutually exclusive categories. Hence, the only statistically important information contained in a categorical variable is the differences in membership. That is, a household placed in one category of the variable is assumed to be distinct on that variable from a household placed in a different category and households placed in the same category are assumed to be the same on that variable.

The recommended procedure for handling categorical variables in multiple regression is to create a set of coded vectors in which households in a given category are assigned an arbitrary number (say 1) while all others not belonging to this category are assigned another arbitrary score (say 0 or -1). In order to avoid artifically introducing linear dependence (multicollinearity) into the model, the number of coded vectors used to represent categorical differences for a given variable is to equal the number of categories in that variable minus one (or c-1 coded vectors).

In reading the tabled values in the ensuing chapters, the reader will note that only c-1 vectors are reflected for each categorical variable. The authors have labelled those vectors in accordance with the category which was arbitrarily assigned a code score of one in the coded vector. Though the c^{th} category doesn't appear with a separate label in the table, it is always included in the analysis by virtue of always being assigned a score of -1 in the set of c-1 coded vectors used to represent the categorical variable. (The coding scheme used is called effect coding.) If the reader wishes to derive a separate coefficient for the c^{th} category, this can be done by taking the algebraic sum of the unstandardized partial regression coefficients for the c-1 labelled categories and then multiplying this sum by -1.

STRUCTURE OF ANALYSIS AND REPORTING OF RELATIONSHIPS

Analyses of the dependent variables and modeling their relationships with independent variables proceeded in four distinct steps: (1) Both independent and dependent variables being considered in each chapter were described; (2) Bivariate relationships between the dependent variable and all considered independent variables were examined and interpretations of results presented; (3) Using only the independent variables retained from the above screening activity (steps 1 and 2), intermediate models were constructed for the dependent variable with each of the three sets of independent variables—household characteristics, supply attributes, and system properties; and (4) The last analytical step was to combine the retained independent variables into a final regression model to assess the overall import of the set of simultaneous relationships between the dependent variable and the retained independent variables. Figure IX-1 summarizes and illustrates the four analysis stages.

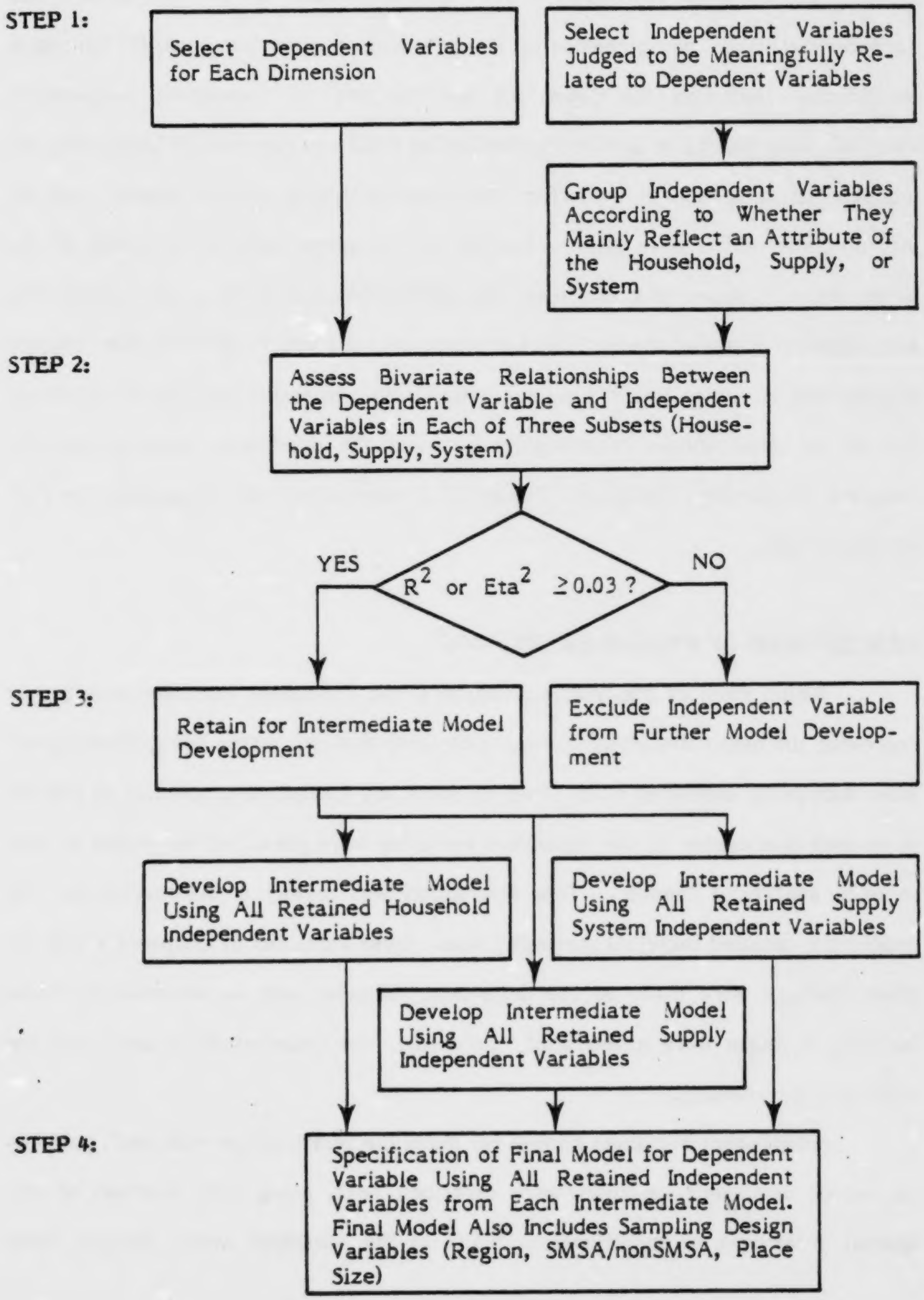
DESCRIPTIONS OF VARIABLES: STEP ONE

Within each of the analytic chapters, the dependent variables selected to represent the major dimension of status of rural drinking water being investigated were succinctly described first. That is, some key statistical properties of the old or revised distribution of the dependent variables were presented in tables or text in each analytical chapter. This description was mainly to summarize for the reader the detailed material presented about these variables in Chapters V and VI. When changes were made to the dependent variable, such as statistically transforming it, these were noted also. Generally, the transformation used was the logarithmic transform.

Independent variables chosen for modeling relationships with each particular set of dependent variables were described next, along with mention of any special treatment or modifications made to the variables which changed their

Figure IX-1

Strategy for Analysis of Each Dependent Variable


STEP 1:
Select Dependent Variables for Each Dimension
Select Independent Variables Judged to be Meaningfully Related to Dependent Variables
Group Independent Variables According to Whether They Mainly Reflect an Attribute of the Household, Supply, or System
STEP 2:
Assess Bivariate Relationships Between the Dependent Variable and Independent Variables in Each of Three Subsets (Household, Supply, System)
YES
R^2 or Eta^2 ≥ 0.03 ?
NO
STEP 3:
Retain for Intermediate Model Development
Exclude Independent Variable from Further Model Development
Develop Intermediate Model Using All Retained Household Independent Variables
Develop Intermediate Model Using All Retained Supply System Independent Variables
Develop Intermediate Model Using All Retained Supply Independent Variables
STEP 4:
Specification of Final Model for Dependent Variable Using All Retained Independent Variables from Each Intermediate Model. Final Model Also Includes Sampling Design Variables (Region, SMSA/nonSMSA, Place Size)

distributions from the way they were previously described in Chapters III through VIII.

BIVARIATE MODELS: STEP TWO

After the variables were introduced, bivariate relationships between each of the candidate independent variables and the dependent variables were computed and examined for strength of association.

As mentioned earlier, the independent variables used consisted of both categorical and continuous household, supply, and system attributes. Categorical variables have a different distributional form than continuous variables. Because of these different forms, they have been analyzed differently, traditionally.

For bivariate models involving two linearly-related, continuously-distributed variables, the measures of association chosen to index the strength of relationship were Pearson's product moment correlation coefficient, r, and the squared value of r, R^2. R^2 is sometimes referred to as the coefficient of determination. It indicates the amount of variance (variability in scores) on the dependent variable which can be known or accounted for with knowledge of the scores of the cases (here, rural households) on the independent variable being regressed (modeled) with the dependent variable. The product-moment coefficient provides a measure of the association between two variables, while the coefficient of determination signifies the proportion of the variance in the dependent variable that may be attributed to the independent variable. Therefore, the product-moment coefficient indicates the magnitude and direction of a relationship, while the coefficient of determination reflects an independent variable's systematic variation.

For bivariate models involving categorical independent variables, however, the relevant statistics are the means and standard deviations on the dependent variable of the entire aggregation of households in a category of the independent variable, and eta^2, which is analogous to the conventional coefficient of determination (R^2). The means and standard deviations are computed to designate group contrasts in terms of the dependent variable, and the eta^2 measures systematic variation in any single attribute's set of categories.

Eta^2 is directly analogous to R^2 and is computed with the same components—sum of squares for the regressors with the dependent variable in the numerator, and total sum of squares of the dependent variable, corrected for the mean, in the denominator. Hence, although the degree of association in these circumstances could have just as correctly been labeled R^2 rather than eta^2, the historical distinction was retained as a reminder to the reader of the difference in the data structure of the independent variable being modeled.

Both R^2 and eta^2 can assume values only between 0 and +1.0. Within these bounds, clearly the closer the value is to 1.0, the stronger the bivariate (or multivariate) relationship is, while the closer the value is to zero, the weaker the relationship. A value of zero indicates no systematic association (correlation) while a value of +1.0 indicates a perfect association.

Because of the large sample size involved in this study, the degrees of freedom associated with the residual sum of squares were very large, rendering the residual mean square very small. Hence a decision was made to forego using conventional statistical modes of evaluating significant relationships (the F test) between independent variables and the dependent variables. Given the extremely low residual mean square, traditionally computed F ratios were always large and always significant in these investigations even when the strength of the relationship was nearly zero, simply because of the small denominator (the residual mean square) used in the F ratio. Instead of using the F test as the decision rule, it was decided to focus on the measures of association directly. Unfortunately, statistical theory or practice has not developed clear standards or guidelines for making decisions about significance of R^2 and eta^2, so the authors postulated their own.

Choice of any value for R^2 and eta^2 was arbitrary. One standard would suggest not retaining any independent variable which could not account for at least 5 percent of the variance in the dependent variable. This would correspond to an R^2 or eta^2 value of 0.05 ($r = 0.25$). However, given the exploratory nature of the whole analysis, the authors decided it was more important to guard against arbitrarily excluding a theoretically or programmatically relevant variable at this stage of analysis (as opposed to including no irrelevant variable) and thereby lowered the decision point to an R^2 or eta^2 value of 0.03.

In the bivariate analysis step, from the original collection of candidate independent variables in each set—household, supply, and supply system—only those which demonstrated an R^2 or eta^2 value of at least 0.03 with the dependent variable were retained for further modeling. These variables have accordingly been designated in the bivariate tables with an underline. Those independent variables which did not have a bivariate R^2 greater than or equal to 0.03 were excluded from further consideration, being judged as "noise" or a random pattern of association. So, based on the strength of the bivariate relationships, some subset of independent variables was retained for further elaboration in the analysis of the dependent variables. The next step of this elaboration was to construct intermediate models.

INTERMEDIATE MODELS OF RELATIONSHIPS: STEP THREE

Whereas the three sets of bivariate relations were undertaken to obtain an initial indication of the strength of patterned covariation, another obvious outcome of this exercise was the selection of variables for further analysis. The prime purpose of the intermediate models was to provide a closer scrutiny of these relationships in the context of other independent variables. Previous research experience has shown that what may appear to be a strong or moderate relationship between two variables based on the bivariate measure of association often becomes attenuated, sometimes drastically altered, when studied in the context of other variables to which both may be associated. Technically, the enterprise involves a more correct specification of the degree of association unconfounded by the effects of other variables which may spuriously produce an apparent relationship, which indeed does not exist when these other relationships are controlled (statistically adjusted). To illustrate, in examining the bivariate relationship between accessibility of supply and mode of conveyance of the water to the household, one observes a moderately strong relationship ($\text{eta}^2 = .44$). However, when these variables are examined in the context of the type of supply variable, the original relationship becomes diminished in importance because of the considerable overlap or relationship between type of supply and mode of conveyance.

The intermediate models attempt to show the interrelationships between the variables in each set and their combined association with the dependent variable. Each model contains a single dependent variable, and two or more independent variables (those categorical or continuous household, supply, or system attributes retained from the bivariate analyses). All models developed in this section are conventional linear regression models of the following general form:

$$Y'_i = a + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \dots\, b_kX_{ik}$$

where

Y'_i = the expected value of the dependent variable
X_{ij} = the explanatory or independent variables
a = the intercept
b_j = the partial regression coefficients.

Models with this structure are evaluated in terms of the intercept and regression coefficients which must be estimated from the pertinent data. Therefore, the intermediate models are routinely accompanied by information on these parameters and other supplementary material.

Generally, these coefficients denote particular aspects of a relationship depending upon whether the corresponding independent variable is categorical or continuous. The coefficients for each continuous variable in the model express the increment in the dependent variable associated with a unit change in the independent variable, controlling for the effects of all other explanatory factors. On the other hand, the coefficients for the categorical independent variables reflect deviations between the unweighted group mean of the dependent variable and the mean of each category after adjusting for the influence of the other independent variables in the model. The unweighted group mean or intercept is considered a composite of the mean scores on all groups that are incorporated into the model. As with the regression coefficients, the intercept is adjusted for the entire set of independent variables.

In addition to examining the behavior of specific bivariate relations in the context of other independent variables, the intermediate models yield a measure of the strength of relationship between the dependent variable and the set of independent variables included in the intermediate model. That measure is again symbolized by R^2, but in the multivariate context has enlarged meaning—the amount of variance in the dependent variable, which can be accounted for by knowledge of scores on the set of independent variables considered simultaneously. By comparing the R^2 values between the three intermediate models—household, supply, and supply system—one can correctly deduce the relative importance of the locus of predominant influence on the dependent variable of concern in the investigation.

Before proceeding to step four, one additional comment about the development of the intermediate models is relevant. These models do not always contain the same independent variables as one moves across dependent variables. Again, the decision rule for retention of a variable for further examination in an intermediate model or the final regression model was the strength of the bivariate relationship between the independent variable and the dependent variable. Clearly, one would not expect the same set of relationships to exist because of the very different nature of the phenomenon encapsulated by the various dependent variables. Hence, each dependent variable was modeled separately at the intermediate analysis stage and the final model stage.

THE FINAL REGRESSION MODEL: STEP FOUR

The last step in the process was to combine all the independent variables used in the intermediate models into a final, overall regression model. The purpose of this step was to reexamine and reevaluate the strength of the relationships in the larger context of all three major sets of influence (independent variables).

In this final step, one additional set of independent variables was introduced beyond those retained from the attributes of the household, supply, and supply system; the sampling design variables were also included on the right hand side of the regression equation, as they carried some important information by virtue of the fact that they determined the very households selected for observation in the first instance. They were not included to claim that the region, SMSA/nonSMSA, and size of place in which the rural household and occupants were located, proved to be universally important in accounting for or predicting variation in water quality, quantity, availability, or cost. They were included to discover if these factors remained important under conditions in which other, more specific characteristics of the household, supply, and system were allowed to operate.

As in the bivariate and intermediate modeling instances, so here also, the overall strength of relationship between the dependent variable being studied and the set of independent variables adjusted for one another and considered simultaneously is indexed by R^2. As before, the closer this value is to zero, the less importance the set of independent variables attain in accounting for variance in the dependent variable, and vice versa when R^2 approaches +1.0. (As a general guideline for evaluation, it might be indicated that in exploratory modeling of relationships from survey data, the R^2 value seldom exceeds 0.40.)

The reader might also note in the ensuing analyses that the R^2 value for the intermediate and final models are often less than the sum of the bivariate R^2 or eta^2 values. This is a function of the independent variables being nonorthogonal or correlated (confounded). It should also be noted that the strength of the relationships between the dependent variables and any independent variable is more accurately reflected by its coefficient in the final model, as it is the more correctly specified model, than the values indicated in the bivariate or intermediate models.

SUMMARY

The purpose of Chapter IX was to introduce the analytical segment of the NSA report. Each of the dimensions of status of rural drinking water was introduced along with the specific measures used in the analysis of each of these dimensions. A rationale was given to support the use of regression analysis for identifying independent variables which had an association with each of the dimensions of rural water supply. Finally, the structure of analysis was detailed.

Several constraints and desired objectives for the analysis dictated the selection of multiple regression as the common analytical framework. Some of these constraints were: (1) wide range of content of the dependent variables, (2) complex phenomena encapsulated in the dependent variables, (3) irregularity of the distributions of some of the dependent variables, (4) use of several independent variables in modeling, (5) use of both categorical and continuous independent variables, (6) unequal number of cases in various categories of the independent variables, and (7) need for a standard, common analytical framework to ease communication of results.

As to the structure of analysis, independent variables were selected based on their expected relationships with the dependent variable under investigation. Bivariate relationships between the candidate independent variables and dependent variable were examined and interpretations made. Only those variables which met or exceeded the reference value used to determine analytical significance ($R^2 \geq 0.03$) were retained for intermediate model development. Intermediate models were constructed for the dependent variable with each of the three sets of independent variables—household characteristics, supply attributes, and system properties. The last stage of the analysis was to develop a final overall regression model which incorporated all the variables contained in the intermediate models and the variables used in the design of the NSA sampling frame (region, SMSA/non-SMSA, size of place).

X

Factors Related to Rural Water Conditions: Quality

Chapter V described components of water quality as it was measured by the NSA throughout rural America. Concentrations of individual substances in rural domestic water supplies and rural residents' perceptions of water quality were reported.

The topics of water quality and rural residents' perceptions of water quality are once again taken up in this chapter. The emphasis shifts from the descriptive (Chapter V) to the analytic in order to present those factors which were apparently related to the quality of household water. Attributes of rural households, the characteristics of their water supplies, and the social organizational aspects of water systems are presented in view of their probable association with the quality of the household water.

Congress, in the Safe Drinking Water Act of 1974, specifically called for an examination of the quality of rural household water. The concept "water quality" is ambiguous and requires definition. Each of the individual contaminants reported in Chapter V was included in the NSA either because of its common occurrence in household water supplies or because of the health hazard it posed. However, addressing the question of general water quality requires an overview that cannot be attained by considering each individual component of water quality separately. Undoubtedly, an investigation of correlates of individual contaminants would be worthwhile, but the NSA is a broad examination of rural water conditions and requires a more compact approach. Composite indicators of water quality, which combine the measures of individual contaminants, are useful for developing an overview. Chapter VI described four composites which will be used in this chapter.

Some contaminants did not lend themselves to the composite approach. In part, this was due to the subsampling done in the NSA. To combine into a composite some contaminants which were measured at every household and some measured on only 10 percent of the households sampled in the NSA would have required that 90 percent of the information on the full-sample contaminants be discarded. In order to avoid that loss, full-sample contaminants were kept apart for portions of the analysis.

Eight indicators of water quality were selected for analysis: total coliform, turbidity, nitrate-N, lead, the aesthetic/economic composite, the health-related composite, the general water quality composite, and the perceived water quality composite. These are designated as dependent variables here and are modeled separately with household attributes, supply characteristics, and system properties.

FULL-SAMPLE INDICATORS

Constituents measured at every household were divided into those with primary MCLs and those with secondary MCLs. There were four contaminants for which the EPA established primary maximum contaminant levels (MCLs) which were measured for every household the NSA studied: total coliform, nitrates, lead, and turbidity. Coliforms are living organisms. They are used to indicate bacterial contamination, the most frequent cause of waterborne health problems. Because of the qualitative difference from the other contaminants and the dominating nature of the potential bacterial health hazard, the count of total coliforms was kept apart for separate analysis. Turbidity is unlike the other contaminants measured in that it is a broad indicator and is not element-specific; to include it in combination with other elements would amount to a form of double-counting. Therefore turbidity, like total coliform, was selected for individual analysis. Keeping the coliform count and the turbidity measure as separate indicators of household water quality left lead and nitrates as the remaining constituents with primary MCLs which were measured at every household. Since there was no intuitively apparent reason nor much economy in presenting them jointly, they also were analyzed separately. The remaining secondary MCL constituents were combined into a composite measure intended to reflect water quality from the standpoint of the contaminants' economic and aesthetic influence. The final full-sample indicator of water quality was the perceived water quality composite. It was derived from the water quality evaluations provided by rural residents. Table X-1 provides a short verbal description of these full-sample water quality measures. Also, pertinent summary statistics are provided in Table X-2.

SUBSAMPLE INDICATORS

There were two subsample composites used in the NSA analysis. The first was formed by combining the subsampled primary MCL-regulated constituents into a health-related index. The other composite combined all of the constituents, subsample and full-sample, into a general indicator of water quality.

Table X-1 provides a short verbal description of these subsample dependent variables. Table X-2 gives the mean and standard deviation for each of the two distributions.

STRUCTURE OF ANALYSIS

The correlates of water quality are first examined using the four individual parameters: total coliform, turbidity, nitrate-N, and lead. The second covers the

Table X-1

Descriptions for Indicators of Household Water Quality

Household Water Quality Indicators	Description
Total coliform	Count of total coliform organisms per 100 milliliters in household water specimens
Nitrate-N	Measured level of nitrate-nitrogen in the household water
Lead	Measured level of lead in the household water
Turbidity	Measured level of turbidity in household water
Aesthetic/economic indicator (Index 2 in Chapter VI)	Standardized combination of sulfates, iron, manganese, color, and total dissolved solids
Healthfulness indicator (Index 1 in Chapter VI)	Combination of the primary MCL-regulated contaminants: arsenic, barium, cadmium, chromium, mercury, selenium, silver, fluoride, gross alpha radiation, lindane, and methoxychlor.
General water quality indicator (Index 3 in Chapter VI)	A gross measure of general water quality with standard scored contributions from total coliform, nitrate-N, sulfates, hardness, iron, manganese, sodium, lead, turbidity, color, total dissolved solids, arsenic, barium, cadmium, chromium, mercury, selenium, silver, fluoride, gross alpha radiation, gross beta radiation, lindane, and methoxychlor
Perceived water quality indicator (Perceived water quality index in Chapter VI)	A composite of household resident's evaluations of the taste, odor, color, sediment, and clarity of the water supply

Table X-2

Means and Standard Deviations for Each of the Indicators of Water Quality

Variable	Mean	Std. Dev.
Total Coliform	10474.6 per 100 milliliters	100,885.7
Nitrate-N	1.7 milligrams per liter	4.5
Lead	0.03 milligrams per liter	0.05
Turbidity	1.5 NTU	6.1
Aesthetic/economic indicator	2.1	2.9
Healthfulness indicator	1.9	2.1
General water quality indicator	9.0	7.3
Perceived water quality indicator	3.3	3.8

correlates of water quality as it was described by the four composites: aesthetic/economic, health-related, general, and perceived. For each of the eight discussions, attention is first given to the sociodemographic characteristics of rural households such as income, education, and ethnicity. Next, the physical facilities of the water supply for extracting, processing, and transporting water is considered. Finally, social-organizational aspects of the system are examined for associations with the quality of the water. For each discussion, bivariate relationships with the water quality indicator are presented. Those which meet an arbitrary minimum screening level (an R^2 or an eta^2 of 0.03) within each set are combined separately in intermediate regression models with the water quality indicator as the dependent variable. Following the three examinations for household, supply, and system variables, a final regression model for that water quality indicator is presented. In addition to the household, supply, and system variables, factors relating to the survey design (region, SMSA/nonSMSA, and size of place) are routinely incorporated in the final model.

Following the analysis of quality for all rural households, a separate presentation is provided which considers only those households served by wells which had fourteen or fewer connections. The same presentation format and analytical techniques are used. Subsequently, two additional analytical presentations are included which consider total coliform only. First, all rural households are assessed; second, those households served by wells with fourteen or fewer connections are examined. Contingency tables are used for these latter sections. In total, the analysis of water quality will have four major sections:

> Water Quality Among all Rural Households: A Regression Approach;
>
> Water Quality Among Rural Households Served by Wells: A Regression Approach;
>
> Total Coliform Among all Rural Households: A Contingency Table Approach;

Total Coliform Among Rural Households Served by Wells: A Contingency Table Approach.

The overall intent of this chapter is to explore the possibility of national patterns of association between either the physical character of the water or rural residents' evaluations of it and the household, supply, and system characteristics. The independent variables used in the analysis are shown in Table X-3.

WATER QUALITY AMONG ALL RURAL HOUSEHOLDS: A REGRESSION APPROACH

Total Coliform

Water can serve as a conducive medium for bacterial growth. There are numerous varieties of microbial forms which are commonly found in water. Those which have attracted the greatest attention historically have been pathogens which originate in the fecal material of infected animals. The coliform group, while not always of fecal origin, and not pathogenic, is used to indicate the level of microbial contamination. Coliforms are generally much more numerous than the pathogens and are more easily detected. Though pathogens are the real concern, if there are viable coliforms in a water supply, then the supply is not fully protected from pathogens. Coliform presence indicates the possibility for pathogens to be transmitted to water consumers. Higher coliform counts generally represent increasing health risk.[1] A fuller discussion of bacterial contamination of rural US water supplies is available in Chapter V.

Each household in the NSA provided a water sample which was examined in a microbiology laboratory for total coliform contamination. The effort here is to

Table X-3

Independent Variables Used in the Analysis of Water Quality

Continuous Variables	Categorical Variables
Household Attributes	
INCOME PEOPLE EDUCATION AGE RESIDENCE UNITS	WORKING (full time, part time, keeping house, retired, unemployed) ETHNICITY (white, nonwhite) TENURE (own, rent, other) MOBILE HOME (mobile home, not mobile home) PERCEIVED COST (no cost, moderate, expensive) WILLINGNESS TO PAY (less, same, more)
Supply Attributes	
TOTAL TREATMENT RELIABILITY ACCESSIBILITY	SUPPLY TYPE (community, well, other)
System Attributes	
CONNECTIONS SERVICE YEARS	OWNERSHIP (public, private individuals and groups, private firm) WATER TESTING (no testing, testing by the household, testing by the system, or by both system and household)

see if there are characteristics of the households, the water supplies, or the systems which appear to be systematically related to the measured level of coliform contamination. Because of the high range of coliform readings and the preponderance of low values, the logarithm of the total coliform count was used for the analysis. The logarithmic values more closely approximate the linear relationships assumption of multiple regression.

BIVARIATE RELATIONSHIPS

Household attributes

A variety of household attributes were assessed in the NSA. Some reflected characteristics of the people, such as the age and education of the household head. Others reflected physical characteristics of the dwelling, such as the number of dwelling units in the structure. Among the characteristics examined were some which may not have a direct link to water quality, but which could reflect complex chains of relationship. For instance, income is probably not directly related to water quality. If a person gets a raise, there is no reason to believe that there will automatically be a change in water quality, but the amount of money available to a household may determine whether or not water treatment devices could be installed, and thus indirectly affect the water quality.

Tables X-4 and X-5 display the degree of association each of the household characteristics has with the log of the measured levels of coliform contamination. For each variable a statistic is given—an eta^2 for categorical variables, the correlation coefficient (r) and the squared correlation coefficient (R^2) for continuous variables. Both of the squared statistics have the same interpretation in the case relating one dependent variable (log coliform count) to one independent (household characteristic) variable. They represent the percentage of variability in the dependent variable which is accounted for with knowledge of the values of the independent variable.

Table X-4

Correlations Between Total Coliform and Continuous Household Variables

Variable	r	R^2
INCOME	-0.12	0.01
PEOPLE	0.03	0.00
EDUCATION	-0.16	0.03*
AGE	0.01	0.00
RESIDENCE	0.06	0.00
UNITS	-0.05	0.00

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-5

Summary Statistics for Total Coliform According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.3	10850.0	102908.0	
Part time	4.8	24968.1	155197.3	
Keeping house	4.8	9609.3	96780.1	
Retired	21.3	6747.2	79934.2	
Unemployed	2.9	9509.3	95332.2	
ETHNICITY				0.00
White	91.4	9970.6	98407.1	
Nonwhite	8.6	15804.2	123980.0	
TENURE				0.00
Own	79.4	10095.0	99321.7	
Rent	17.8	10527.2	100497.3	
Other	2.7	21211.6	140320.5	
MOBILE HOME				0.00
Not mobile home	87.7	9821.4	97597.7	
Mobile home	12.3	15137.8	121630.0	
PERCEIVED COST				0.04*
No cost	4.5	69418.7	252846.8	
Moderate	83.8	6990.8	82151.8	
Expensive	11.7	15812.4	124692.7	
WILLINGNESS TO PAY				0.01
Less	3.6	7483.4	85406.1	
Same	63.4	6932.9	82365.6	
More	33.0	17577.8	130027.2	

N = 21,974,000
n = 2,654

---

*Underscoring denotes that the variable was retained for further analysis.

It is clear from Tables X-4 and X-5 that characteristics of the rural household were not strongly related to coliform counts. Among the variables considered, only four registered a correlation above zero. The variables retained were perceived cost and the education of the head of the household. Households which reported that their water was obtained at no cost showed the highest average coliform counts. Households which said their water was expensive had lower counts, on the average, than the 'no cost' households. Most households—almost 84 percent—thought their water costs were about right. This group had the lowest average number of coliforms detected. Education of the household head showed a weak inverse relationship with coliform counts, indicating a slight tendency of higher coliform counts among households with lower education levels.

Supply attributes

The NSA used the term "supply" to mean the set of facilities for extracting, treating, and transporting water to the household (see Chapter II).

There were four variables which were measured on all types of supplies throughout rural America. The first variable was the classification of the supply as either community water system, well or other (surface water, springs, cisterns, hauled, and purchased bottled supplies). The second variable was a reliability index based upon the frequency and severity of breakdowns at the household (see Chapter VI). The third variable was a composite measure of accessibility. The accessibility composite was intended to reflect how much effort the household members had to expend in order to obtain water: for most community systems, the household simply connects to a water main; for a well, typically, one must install a pump, a pressure tank, possibly some treatment devices, and so forth. Wells, therefore, were generally assigned lower accessibility scores than large multiple connection supplies (see Chapter VI). The fourth variable was a count of the types of devices used to treat or handle the water, both at the system and at the household. (For cases where the supply was on the premises, the count was, of course, only for the household.)

Tables X-6 and X-7 indicate that of the four supply variables studied, the type of supply was the most useful in understanding water quality. This bivariate relation showed that households served by community water supplies generally had lower coliform counts than those served by wells. Both community supplies and wells had lower average coliform counts than households served by other supplies.

Measures of accessibility and the number of water treatment devices both showed weak relationships with coliform counts. Less accessible supplies and supplies using fewer devices had a slight tendency toward lower coliform counts. No association was apparent for the reliability indicator.

- Supply type and accessibility were retained for examination in an intermediate regression model.

System attributes

The use of the term "system" in the NSA is for specifying the social-organizational aspects of supplies (see Chapter II). The NSA measured four descriptors for all types of systems: number of service connections, system ownership, the number of years rural dwellers had been served by the system, and an indication of whether any water quality testing had been done. The number of years served (SERVICE YEARS) was the age of the system, or time since the last major alteration, for multiple connection systems; for single connection systems, it was the number of years householders reported being served by the system. The number of connections ranged from one through tens of thousands. A logarithmic transformation of the number of connections made the variable more compatible with the assumptions underlying the regression model.

Tables X-8 and X-9 show that three of the system variables have an R^2 or eta^2 above 0.03 with the log of coliform counts. The number of years the residents

Table X-6

Correlations Between Total Coliform and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.13	0.02
RELIABILITY	0.00	0.00
ACCESSIBILITY	0.27	0.07*

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-7

Summary Statistics for Total Coliform According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.16*
Community	50.0	3437.9	38286.7	
Well	44.8	10606.8	101227.0	
Other	5.2	76951.5	264264.0	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-8

Correlations Between Total Coliform and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.30	0.09*
SERVICE YEARS	0.09	0.01

N = 21,974,000
n = 2,654

---

*Underscoring denotes that the variable was retained for further analysis.

Table X-9

Summary Statistics for Total Coliform According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.08*
Private groups and individuals	52.3	15688.3	122954.0	
Public	40.7	3435.3	58171.7	
Private firm	6.7	8765.8	93075.6	
Other	0.3	83820.0	277079.0	
WATER TESTING				0.09
No testing	29.8	23720.0	150648.0	
Household tests	1.4	9742.0	97312.8	
System tests, or both household and system tests	52.9	3256.8	56672.1	

N = 21,974,000
n = 2,654

---

*Underscoring denotes that the variable was retained for further analysis.

had been served by the supply apparently was not related to the degree of coliform contamination. The correlation between the number of connections and the coliform count suggested that larger systems tended to have fewer coliforms in the water. Similarly, the average coliform count among publicly owned systems was lower than for other types of ownership. It usually was these larger, publicly owned systems which tested the water. Indeed, Table X-9 shows that household supplies which were tested by the system had a lower likelihood of coliform presence.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

Table X-10 presents the combined effects of considering perceived cost and education. Together, these two independent variables accounted for only 7 percent of the variability in total coliform counts. In general, then, the degree of coliform contamination in rural America was not very strongly related to any particular household characteristic or group of characteristics. Given the preponderance of one or no coliforms per 100 milliliters of water (72 percent), this finding was not unexpected, however.

Intermediate supply model

Table X-11 indicates that 17 percent of the variability in the log of coliform counts was accounted for with knowledge of the type of supply and the accessibility score. Recall, however, from Table X-7, that the type of supply alone explained 16 percent of coliform's variability. The association shown in Table X-11 between coliform counts and the combination of variables is largely a function of supply type. Community supplies tended to be more accessible than other supplies. Furthermore, community supplies tended to have more water treatment devices

Table X-10

Intermediate Regression Estimates for Total Coliform and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	2.740
	PERCEIVED COST		
PC_1	No cost	b_1	1.508
PC_2	Moderate	b_2	-0.576
X_1	EDUCATION	b_3	-0.094

$R^2 = 0.07$

Adjusted $R^2 = 0.07$

Standard error of estimate = 2.318

N = 21,974,000
n = 2,654

$$Y' = a + b_1PC_1 + b_2PC_2 + b_3X_1$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-11

Intermediate Regression Estimates for Total Coliform and Supply Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	1.824
	SUPPLY TYPE		
ST_1	Community	b_1	-1.534
ST_2	Well	b_2	-0.478
X_1	ACCESSIBILITY	b_3	0.035

$R^2 = 0.17$

Adjusted $R^2 = 0.17$

Standard error of estimate = 2.162

$N = 21{,}974{,}000$
$n = 2{,}654$

$Y' = a + b_1ST_1 + b_2ST_2 + b_3X_1$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

than smaller supplies. Consequently, the combined model of Table X-11 does not display much improvement over simply knowing the supply type.

Intermediate system model

Table X-12 indicates that the three retained system variables, combined, accounted for 10 percent of the variability in coliform counts. Therefore, among those system attributes measured across all types of systems, there were no strong linear associations with the log counts of total coliform bacteria.

FINAL MODEL SPECIFICATION

A general model which combines the retained household, supply, and system variables is presented in Table X-13. In addition, variables representing the design effects (region, SMSA, and size of place) were also included. The model showed a R^2 (adjusted for the number of variables) of 0.19. That is, knowledge of these design, household, supply, and system characteristics accounted for only one-fifth of the variability in log coliform counts. Those characteristics which were strongly related to the type of water supply (community, well, or other) tended to account for essentially the same variability in the coliform counts. Hence, when they were combined, there was not a dramatic increase in association beyond that which was shown by the type of supply. Table X-14 illustrates the overlap of the variables in that it contains only a few of the variables of Table X-13, yet still demonstrates an adjusted R^2 of 0.19.

In general, coliform contamination did not show a strong linear association with any of the household, supply, or system characteristics measured. A large proportion (over 70 percent) of rural US households had zero or one coliform identified per 100 milliliters of water. This condition of wide variation on the independent measures, together with an almost constant dependent variable, suppressed the measure of association.

Table X-12

Intermediate Regression Estimates for Total Coliform and System Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	1.944
X_1	CONNECTIONS	b_1	-0.123
	WATER TESTING		
WT_1	Household tests	b_2	-0.058
WT_2	System tests, or both household and system tests	b_3	-0.291
	OWNERSHIP		
O_1	Public	b_4	-0.179
O_2	Private groups and individuals	b_5	-0.132
O_3	Private firm	b_6	-0.281

$R^2 = 0.10$

Adjusted $R^2 = 0.10$

Standard error of estimate = 2.284

N = 21,974,000
n = 2,654

$$Y' = a + b_1X_1 + b_2WT_1 + b_3WT_2 + b_4O_1 + b_5O_2 + b_6O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-13

Final Regression Estimates for Total Coliform

Variable Designation		Parameter	Estimate
	Intercept	a	2.665
	REGION		
R_1	Northeast	b_1	-0.007
R_2	North Central	b_2	-0.417
R_3	South	b_3	0.228
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.239
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	0.026
SI_2	Small rural communities	b_6	-0.050
	SUPPLY TYPE		
ST_1	Community	b_7	-0.810
ST_2	Well	b_8	-0.741
X_1	CONNECTIONS	b_9	-0.075
	WATER TESTING		
WT_1	Household tests	b_{10}	0.136
WT_2	System tests, or both household and system tests	b_{11}	-0.264
	OWNERSHIP		
O_1	Public	b_{12}	0.012
O_2	Private groups and individuals	b_{13}	0.056
O_3	Private firm	b_{14}	-0.125
X_2	ACCESSIBILITY	b_{15}	0.024

Table X-13 continued

Variable Designation		Parameter	Estimate
	PERCEIVED COST		
PC_1	No cost	b_{16}	0.049
PC_2	Moderate	b_{17}	-0.169
X_3	EDUCATION	b_{18}	-0.039

$R^2 = 0.20$

Adjusted $R^2 = 0.19$

Standard error of estimate = 2.152

N = 21,974,000
n = 2,654

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7ST_1 + b_8ST_2 + b_9X_1 + b_{10}WT_1 + b_{11}WT_2 + b_{12}O_1 + b_{13}O_2 + b_{14}O_3 + b_{15}X_2 + b_{16}PC_1 + b_{17}PC_2 + b_{18}X_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-14

Final Regression Estimates for Total Coliform Using a Reduced Number of Variables

Variable Designation		Parameter	Estimate
	Intercept	a	2.959
	REGION		
R_1	Northeast	b_1	-0.019
R_2	North Central	b_2	-0.371
R_3	South	b_3	0.225
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.314
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	0.014
SI_2	Small rural communities	b_6	0.001
	SUPPLY TYPE		
ST_1	Community	b_7	-1.563
ST_2	Well	b_8	-0.522
	PERCEIVED COST		
PC_1	No cost	b_9	0.589
PC_2	Moderate	b_{10}	-0.204
X_1	EDUCATION	b_{11}	-0.047

$R^2 = 0.19$

Adjusted $R^2 = 0.19$

Standard error of estimate = 2.162

N = 21,974,000
n = 2,654

Table X-14 continued

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7ST_1 + b_8ST_2 + b_9PC_1 + b_{10}PC_2 + b_{11}X_1$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Turbidity

Turbidity, a measure of the suspended solids in water, does not refer to any specific constituent, but to a general physical condition of water. Consequently, it represents a potential redundancy when measured along with other specific constituents.

Despite its lack of specificity, turbidity can be an important indicator of water quality. The removal of solids from water can eliminate the risks from a wide variety of contaminants. Apart from the elemental nature of the particles, hazards may also occur from absorbed materials (such as polycyclic aromatic hydrocarbons), adsorbed or adhering materials (such as bacteria), and chemical reactions (such as the formation of trihalomethanes from chlorination).[2] The removal of suspended particles improves water quality directly by eliminating any direct effect of the particles, and indirectly by reducing interference with disinfection or by reducing solids which could serve as vehicles for concentrating, transporting, or releasing other pollutants.[3]

BIVARIATE RELATIONSHIPS

Household attributes

Tables X-15 and X-16 indicate that the incidence of turbid household water supplies was not systematically associated with characteristics of the households themselves. The R^2 values for the household characteristics rarely exceeded zero and never exceeded 0.02, a value below the minimum screening-level criterion.

Supply attributes

Tables X-17 and X-18 indicate that the only supply variable which passed the minimum association level was the type of supply. The mean turbidity levels for type of supply suggest that wells tended to be more susceptible to turbidity

Table X-15

Correlations Between Turbidity and Continuous Household Variables

Variable	r	R^2
INCOME	-0.09	0.01
PEOPLE	-0.04	0.00
EDUCATION	-0.13	0.02
AGE	0.07	0.00
RESIDENCE	0.08	0.01
UNITS	-0.04	0.00

N = 21,974,000
n = 2,654

Table X-16

Summary Statistics for Turbidity According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.3	1.25	4.70	
Part time	4.8	1.77	9.11	
Keeping house	4.8	1.74	5.33	
Retired	21.3	1.97	8.58	
Unemployed	2.9	1.52	7.23	
ETHNICITY				0.00
White	91.4	1.41	5.86	
Nonwhite	8.6	1.98	7.99	
TENURE				0.00
Own	79.4	1.53	6.25	
Rent	17.8	0.96	4.55	
Other	2.7	2.50	8.78	
MOBILE HOME				0.00
Not mobile home	87.7	1.42	6.04	
Mobile home	12.3	1.68	6.30	
PERCEIVED COST				0.00
No cost	4.5	2.12	7.73	
Moderate	83.8	1.51	6.31	
Expensive	11.7	0.96	4.17	
WILLINGNESS TO PAY				0.00
Less	3.6	1.16	3.73	
Same	63.4	1.34	5.95	
More	33.0	1.72	6.49	

N = 21,974,000
n = 2,654

Table X-17

Correlations Between Turbidity and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.06	0.00
RELIABILITY	0.02	0.00
ACCESSIBILITY	0.08	0.01

N = 21,974,000
n = 2,654

Table X-18

Summary Statistics for Turbidity According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.04*
Community	50.0	0.61	2.76	
Well	44.8	2.46	8.47	
Other	5.2	0.54	0.44	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

than community supplies or other types of supply. Most wells were relatively free of turbidity—the differences in means were a result of a few wells which produced highly turbid water. Eta^2 remained low, even though the mean differences appeared to be fairly substantial, because of the large variability among scores in each type of supply. The large range of turbidity values was reflected in the size of the standard deviation of turbidity levels for wells.

These findings should not be interpreted as a reflection on the relative turbidity of source water. Groundwater generally is less turbid than surface water. But community water supplies, which are the most likely type of supply to rely upon surface water, tend to be engineered with turbidity reduction in mind. Individual household wells, on the other hand, usually pass the source water on to the consumer without much treatment. So, when they are both still in the source, groundwater may tend to be less turbid than surface water. But whatever turbidity is present in the water extracted by individual household wells is generally passed on to the users. These findings indicate that regardless of source, most community systems were supplying relatively nonturbid water.

System attributes

Among those attributes of systems measured in the NSA, three had levels of association sufficient to merit inclusion in the intermediate regression model (see Tables X-19 and X-20). The logarithm of the number of connections, system ownership, and history of water testing each showed a 0.04 measure of association with the logarithm of turbidity. Systems with few connections had a slight tendency toward turbidity, which was consistent with the earlier indication that individual wells had higher turbidity than community systems. Similarly, privately owned systems, usually individual wells, showed the highest average turbidity levels—more than three times higher than the average turbidity level for any of the other categories of ownership. Furthermore, households for which no water

Table X-19

Correlations Between Turbidity and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.21	0.04*
SERVICE YEARS	0.05	0.00

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-20

Summary Statistics for Turbidity According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				
Private groups and individuals	52.3	2.28	8.23	0.04*
Public	40.7	0.54	1.34	
Private firm	6.7	0.64	1.22	
Other	0.3	0.54	0.44	
WATER TESTING				
No testing	29.8	2.46	8.93	0.04
Household tests	17.3	2.35	7.89	
System tests, or both household and system tests	52.9	0.60	1.72	

N = 21,974,000
n = 2,654

---

*Underscoring denotes that the variable was retained for further analysis.

testing was reported and those which tested at the household showed the highest average turbidity values. Households in these two groups tended to have individual household supplies—usually wells as the major supply of water.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

None of the household variables had a sufficiently strong association with the measure of turbidity to be included in an intermediate model.

Intermediate supply model

Only one variable, the type of supply, had an association above 0.03 with the measure of turbidity. An intermediate regression model is omitted because it would only replicate the association shown in Table X-18.

Intermediate system model

Table X-21 displays the results of combining in an intermediate regression model the effects of three system variables: number of connections, water testing, and system ownership. Together they demonstrated an association of 0.04 with turbidity levels. Their combined effect was no larger than their separate associations with turbidity.

FINAL MODEL SPECIFICATION

Including supply type along with the system characteristics in a final model which also contained the design characteristics did not add substantially to the information already presented. Table X-22 shows an adjusted R^2 of 0.06. Supply and system characteristics together accounted for about 5 percent of the variability; the design factors contributed about 1 percent.

Table X-21

Intermediate Regression Estimates for Turbidity and System Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	0.510
X_1	CONNECTIONS	b_1	-0.016
	WATER TESTING		
WT_1	Household tests	b_2	0.025
WT_2	System tests, or both household and system tests	b_3	-0.071
	OWNERSHIP		
O_1	Public	b_4	0.029
O_2	Private groups and individuals	b_5	0.067
O_3	Private firm	b_6	0.035

$R^2 = 0.05$

Adjusted $R^2 = 0.04$

Standard error of estimate = 0.599

N = 21,974,000
n = 2,654

$$Y' = a + b_1X_1 + b_2WT_1 + b_3WT_2 + b_4O_1 + b_5O_2 + b_6O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-22

Final Regression Estimates for Turbidity

Variable Designation		Parameter	Estimate
	Intercept	a	0.480
	REGION		
R_1	Northeast	b_1	-0.098
R_2	North Central	b_2	-0.075
R_3	South	b_3	0.031
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.023
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.029
SI_2	Small rural communities	b_6	0.069
	SUPPLY TYPE		
ST_1	Community	b_7	-0.039
ST_2	Well	b_8	0.069
X_1	CONNECTIONS	b_9	-0.004
	WATER TESTING		
WT_1	Household tests	b_{10}	0.019
WT_2	System tests, or both household and system tests	b_{11}	-0.060
	OWNERSHIP		
O_1	Public	b_{12}	0.006
O_2	Private groups and individuals	b_{13}	0.066
O_3	Private firm	b_{14}	0.045

$R^2 = 0.06$

Adjusted $R^2 = 0.06$

Standard error of estimate = 0.595

Table X-22 continued

N = 21,974,000
n = 2,654

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7ST_1 + b_8ST_2 + b_9X_1 + b_{10}WT_1 + b_{11}WT_2 + b_{12}O_1 + b_{13}O_2 + b_{14}O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

The incidence of turbidity was not closely associated with particular household, supply, or system characteristics. While mean differences among various categories of households did show changes in turbidity levels, the range of values across categories was so large and overlapping that no linear association of any strength was demonstrated.

Nitrate-N

Nitrate contamination of water is an important health concern. Nitrates can be converted to nitrites by bacteria in the gastrointestinal tract. The resulting nitrite oxidizes hemoglobin to methemoglobin, which cannot transmit oxygen. Cellular anoxia results. The greatest danger is to infants. Additionally, nitrites can react with amines in consumed food to form nitrosamines which are potentially carcinogenic, teratogenic and mutagenic.[4]

Nitrates enter water supplies in a variety of ways. Common sources are erosion and municipal and industrial wastewaters. Feedlot runoff, sewage disposal systems, and agricultural fertilizer are also nitrate sources, among others.

Chapter V described the distribution of nitrate concentrations in rural water supplies. Over 97 percent of all rural households were found to have levels lower than the NSA reference value of ten milligrams of nitrate-N per milliliter of water.

BIVARIATE RELATIONSHIPS

Household attributes

Tables X-23 and X-24 indicate that none of the household characteristics used in this analysis had a meaningful association with the degree of nitrate concentration. Only the length of residence manifested an R^2 above zero. None of the average nitrate levels shown in Table X-24 showed meaningful differences.

Table X-23

Correlations Between Nitrate-N and Continuous Household Variables

Variable	r	R^2
INCOME	-0.02	0.00
PEOPLE	0.00	0.00
EDUCATION	-0.06	0.00
AGE	0.02	0.00
RESIDENCE	0.09	0.01
UNITS	-0.04	0.00

N = 21,974,000
n = 2,654

Table X-24

Summary Statistics for Nitrate-N According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.3	1.66	4.41	
Part time	4.8	2.39	5.97	
Keeping house	4.8	1.56	3.07	
Retired	21.3	1.67	4.66	
Unemployed	2.9	2.46	4.00	
ETHNICITY				0.00
White	91.4	1.73	4.58	
Nonwhite	8.6	1.46	3.10	
TENURE				0.00
Own	79.4	1.67	4.51	
Rent	17.8	1.69	3.34	
Other	2.7	2.78	8.23	
MOBILE HOME				0.00
Not mobile home	87.7	1.74	4.61	
Mobile home	12.3	1.47	3.33	
PERCEIVED COST				0.00
No cost	4.5	2.27	6.19	
Moderate	83.8	1.77	4.66	
Expensive	11.7	1.22	2.19	
WILLINGNESS TO PAY				0.00
Less	3.6	1.53	3.49	
Same	63.4	1.81	4.69	
More	33.0	1.54	4.13	

N = 21,974,000
n = 2,654

Every category's average nitrate level was closer than one standard deviation to the levels of the other categories. The large standard deviations indicated that each category had a wide range of nitrate readings.

The findings of Tables X-23 and X-24 indicate that the occurrence of nitrates in domestic water supplies was randomly distributed with respect to these indicators of household differences. These findings suggested that the nitrate concentrations of rural domestic water supplies were not systematically associated with household characteristics.

Supply attributes

Among the four supply characteristics studied, only one showed an appreciable association with nitrate levels (Tables X-25 and X-26). The type of supply—whether well, community, or other—was weakly related to the level of nitrates in rural water supplies, accounting for 4 percent of the variability in nitrate levels. It appeared that community water supplies experienced slightly lower nitrate concentrations, on the average, than wells, but the association was not strong. "Other" supplies, which amounted to 5.2 percent of all supplies, had the lowest average nitrate concentrations of the three supply categories.

System attributes

Four characteristics of water systems were measured on every rural system. Of these, three barely met the minimum R^2 or eta^2 value of 0.03.

The logarithm of the number of connections showed an R^2 of 4 percent (Table X-27). System ownership and water testing (Table X-28) also showed weak associations with nitrate readings.

Table X-25

Correlations Between Nitrate-N and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.11	0.01
RELIABILITY	-0.03	0.00
ACCESSIBILITY	0.05	0.00

N = 21,974,000
n = 2,654

Table X-26

Summary Statistics for Nitrate-N According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.04*
Community	50.0	1.09	2.25	
Well	44.8	2.47	6.14	
Other	5.2	0.57	0.53	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-27

Correlations Between Nitrate-N and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.20	0.04*
SERVICE YEARS	-0.12	0.01

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-28

Summary Statistics for Nitrate-N According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.03*
Private groups and individuals	52.3	2.28	5.77	
Public	40.7	1.11	2.22	
Private firm	6.7	0.90	1.80	
Other	0.3	1.11	1.51	
WATER TESTING				0.04
No testing	29.8	2.27	5.13	
Household tests	17.3	2.58	7.23	
System tests, or both household and system tests	52.9	1.10	2.25	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

None of the household attributes demonstrated sufficient association with nitrate levels to be included in an intermediate regression model.

Intermediate supply model

Only one supply variable, supply type, surpassed the R^2 criterion of at least 0.03. An intermediate regression model was not warranted since it would merely repeat the association shown in Table X-26.

Intermediate system model

Three system variables (number of connections, water testing, and system ownership) were included in the regression model of Table X-29. Together, they registered an R^2 of 0.04, no more than each of them contributed separately.

FINAL MODEL SPECIFICATION

Table X-30 shows the combined power of the sample design variables with supply type, the number of connections, water test history, and system ownership to account for the observed levels of nitrate concentrations across the country. Together they accounted for only 7 percent of the variability of nitrate levels. This indicated that across the nation there was wide variation of nitrate levels among systems of every size, type, ownership, and so on. There did not appear to be any systematic, nationwide pattern with nitrate concentrations. In general, nitrate levels had no strong association with the examined household, supply, or system characteristics.

Table X-29

Intermediate Regression Estimates for Nitrate-N and System Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	1.877
X_1	CONNECTIONS	b_1	-0.105
	WATER TESTING		
WT_1	Household tests	b_2	0.367
WT_2	System tests, or both household and system tests	b_3	-0.433
	OWNERSHIP		
O_1	Public	b_4	0.467
O_2	Private groups and individuals	b_5	0.364
O_3	Private firm	b_6	0.002

$R^2 = 0.04$

Adjusted $R^2 = 0.04$

Standard error of estimate = 0.631

N = 21,974,000
n = 2,654

$$Y' = a + b_1X_1 + b_2WT_1 + b_3WT_2 + b_4O_1 + b_5O_2 + b_6O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-30
Final Regression Estimates for Nitrate-N

Variable	Designation	Parameter	Estimate
	Intercept	a	0.725
	REGION		
R_1	Northeast	b_1	-0.005
R_2	North Central	b_2	0.054
R_3	South	b_3	-0.148
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.041
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	0.067
SI_2	Small rural communities	b_6	0.007
	SUPPLY TYPE		
ST_1	Community	b_7	0.004
ST_2	Well	b_8	0.007
X_1	CONNECTIONS	b_9	-0.008
	WATER TESTING		
WT_1	Household tests	b_{10}	0.028
WT_2	System tests, or both household and system tests	b_{11}	-0.091
	OWNERSHIP		
O_1	Public	b_{12}	0.063
O_2	Private groups and individuals	b_{13}	0.069
O_3	Private firm	b_{14}	-0.023

$R^2 = 0.07$

Adjusted $R^2 = 0.07$

Standard error of estimate = 0.621

Table X-30 continued

N = 21,974,000
n = 2,654

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7ST_1 + b_8ST_2 + b_9X_1 + b_{10}WT_1 + b_{11}WT_2 + b_{12}O_1 + b_{13}O_2 + b_{14}O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Lead

Lead is a toxic metal. Even low exposure, if it is prolonged, is dangerous since lead can accumulate in the human body and reach injurious levels. Lead in soldered pipe joints and, in some old supplies, in the pipes themselves, make lead contact with most domestic water supplies unavoidable. Soft, low-pH water generally presents the greatest likelihood of picking up traces of lead while it is in the distribution system.[5]

Chapter V displayed the NSA results for the national situation of lead concentrations in rural water supplies. Over 3.5 million rural households were projected to have lead concentrations in their water supplies above the NSA reference value of 0.05 milligrams per liter of water.

BIVARIATE RELATIONSHIPS

Household attributes

Among the items which were consistently measured at every household, none had an R^2 or an eta^2 above 0.01 (Tables X-31 and X-32). Furthermore, the differences among average lead concentrations for the various categories of Table X-32 are never greater than 0.02 milligrams per liter. It does not appear that there were systematic national patterns which related lead concentrations to household characteristics.

Supply attributes

Tables X-33 and X-34 relate the supply attributes with lead concentrations in household supplies. Here again, no association was apparent. It may be that the construction and condition of the plumbing, in conjunction with a measure of the pH of the water, would show a meaningful association with lead levels. Unfortunately, neither type of information was obtained in the NSA.

Table X-31

Correlations Between Lead and Continuous Household Variables

Variable	r	R^2
INCOME	-0.09	0.01
PEOPLE	-0.02	0.00
EDUCATION	-0.08	0.01
AGE	-0.02	0.00
RESIDENCE	-0.00	0.00
UNITS	-0.03	0.00

N = 21,974,000
n = 2,654

Table X-32

Summary Statistics for Lead According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.3	0.03	0.05	
Part time	4.8	0.02	0.03	
Keeping house	4.8	0.03	0.04	
Retired	21.3	0.04	0.07	
Unemployed	2.9	0.04	0.08	
ETHNICITY				0.00
White	91.4	0.03	0.05	
Nonwhite	8.6	0.03	0.04	
TENURE				0.00
Own	79.4	0.03	0.05	
Rent	17.8	0.03	0.06	
Other	2.7	0.03	0.04	
MOBILE HOME				0.00
Not mobile home	87.7	0.03	0.05	
Mobile home	12.3	0.04	0.06	
PERCEIVED COST				0.00
No cost	4.5	0.04	0.05	
Moderate	83.8	0.03	0.05	
Expensive	11.7	0.03	0.06	
WILLINGNESS TO PAY				0.00
Less	3.6	0.03	0.05	
Same	63.4	0.03	0.05	
More	33.0	0.03	0.06	

N = 21,974,000
n = 2,654

Table X-33

Correlations Between Lead and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.05	0.00
RELIABILITY	0.03	0.00
ACCESSIBILITY	0.05	0.00

N = 21,974,000
n = 2,654

Table X-34

Summary Statistics for Lead According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.00
Community	50.0	0.03	0.06	
Well	44.8	0.03	0.04	
Other	5.2	0.01	0.01	

N = 21,974,000
n = 2,654

System attributes

System characteristics, like both household and supply characteristics, showed no systematic national association with lead concentrations (see Tables X-35 and X-36.)

INTERMEDIATE AND FINAL MODEL SPECIFICATION

The consideration of those household, supply, and system attributes measured in the NSA across all households did not provide useful indicators of lead concentrations. There did not seem to be any nationwide systematic association with measured lead levels. Lead, in low concentrations, seemed to be fairly pervasive. When it occurred in higher concentrations, it apparently did so because of conditions which were not ascertained in the NSA. No intermediate or final regression models were justified, given the uniform lack of association with lead levels demonstrated by household, supply, and system variables.

Aesthetic/Economic Composite

The discussion of household water quality has thus far centered upon individual constituents. The following presentations turn to a consideration of composite measures. A detailed discussion of the construction of the composites is contained in Chapter VI.

The disadvantage of a composite is that individual parameters lose their unique interpretation, but there are at least two important advantages. First, a properly constructed composite is a summary measure which can give an overview that is difficult to achieve by examining the individual components. Second, a composite is able to reflect multiple high or low readings across several constituents at one household. A knowledgeable and careful reader may be able to study

Table X-35

Correlations Between Lead and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.01	0.00
SERVICE YEARS	-0.03	0.00

N = 21,974,000
n = 2,654

Table X-36

Summary Statistics for Lead According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	52.3	0.03	0.05	
Public	40.7	0.03	0.06	
Private firm	6.7	0.03	0.05	
Other	0.3	0.01	0.01	
WATER TESTING				0.00
No testing	29.8	0.03	0.04	
Household tests	17.3	0.03	0.05	
System tests, or both household and system tests	52.9	0.03	0.06	

N = 21,974,000
n = 2,654

results for individual constituents and attain an overview without the aid of a composite, but without a composite, it is impossible to detect when high, middle, or low readings occur in concert at a single household.

The aesthetic/economic composite was constructed from constituents which were measured at every household. They are items which have secondary MCLs. They can affect the aesthetic quality of the water, and may have economic implications for plumbing and for laundry wear and discoloring. A higher composite score implied a greater likelihood that the water supply would have objectionable aesthetic characteristics.

Five items were used to form the aesthetic/economic composite: sulfates, iron, manganese, color, and estimated total dissolved solids. Composite scores could range from 0 to 49, but most households clustered near the bottom. The mean was 2.13. The standard deviation was 2.91.

BIVARIATE RELATIONSHIPS

Household attributes

Tables X-37 and X-38 showed that the composite formed from these secondary MCL items did not vary systematically with any of the household attributes studied.

Supply attributes

Tables X-39 and X-40 showed no systematic association between supply characteristics and the aesthetic/economic composite.

System attributes

Tables X-41 and X-42 did not indicate any notable associations between system attributes and the aesthetic/economic composite. That is, no R^2s or eta^2s were greater than or equal to 0.03.

Table X-37

Correlations Between the Aesthetic/Economic Composite and Continuous Household Variables

Variable	r	R^2
INCOME	-0.02	0.00
PEOPLE	-0.00	0.00
EDUCATION	-0.05	0.00
AGE	0.05	0.00
RESIDENCE	0.01	0.00
UNITS	-0.01	0.00

N = 21,974,000
n = 2,654

Table X-38

Summary Statistics for the Aesthetic/Economic Composite According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.3	2.09	2.94	
Part time	4.8	2.11	2.10	
Keeping house	4.8	2.22	3.14	
Retired	21.3	2.28	3.04	
Unemployed	2.9	1.75	1.80	
ETHNICITY				0.00
White	91.4	2.12	2.93	
Nonwhite	8.6	2.23	2.58	
TENURE				0.00
Own	79.4	2.12	2.86	
Rent	17.8	2.08	2.36	
Other	2.7	3.00	5.83	
MOBILE HOME				0.00
Not mobile home	87.7	2.13	2.97	
Mobile home	12.3	2.18	2.42	
PERCEIVED COST				0.00
No cost	4.5	2.44	3.75	
Moderate	83.8	2.14	2.98	
Expensive	11.7	2.03	2.17	
WILLINGNESS TO PAY				0.01
Less	3.6	1.75	2.02	
Same	63.4	1.91	2.56	
More	33.0	2.61	3.50	

N = 21,974,000
n = 2,654

Table X-39

Correlations Between the Aesthetic/Economic Composite and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.02	0.00
RELIABILITY	-0.05	0.00
ACCESSIBILITY	0.01	0.00

N = 21,974,000
n = 2,654

Table X-40

Summary Statistics for the Aesthetic/Economic Composite According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.01
Community	50.0	1.86	2.09	
Well	44.8	2.51	3.63	
Other	5.2	1.64	1.74	

N = 21,974,000
n = 2,654

Table X-41

Correlations Between the Aesthetic/Economic Composite and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.11	0.01
SERVICE YEARS	0.01	0.00

N = 21,974,000
n = 2,654

Table X-42

Summary Statistics for the Aesthetic/Economic Composite According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.01
Private groups and individuals	52.3	2.41	3.50	
Public	40.7	1.86	2.10	
Private firm	6.7	1.72	1.54	
Other	0.3	1.64	1.73	
WATER TESTING				0.01
No testing	29.8	2.27	3.33	
Household tests	17.3	2.72	3.95	
System tests, or both household and system tests	52.9	1.86	2.10	

N = 21,974,000
n = 2,654

INTERMEDIATE AND FINAL MODEL SPECIFICATION

No intermediate or final regression models were appropriate. There was no indication of any systematic national association between household, supply, or system attributes and the aesthetic/economic character of rural water supplies.

Health-Related Composite

Eleven constituents were combined to form a health-related composite: arsenic, barium, cadmium, chromium, mercury, selenium, silver, fluoride, gross alpha radiation, lindane, and methoxychlor. Chapter VI described the details of the construction of this composite.

This second composite differs from the aesthetic/economic composite in three important respects. Each of the items poses a definite potential health hazard. They all have primary MCLs. Finally, they were measured only at a 10 percent subsample of NSA households.

Chapter V described the general distributional patterns of each item of the health-related composite. While many of these components can be extremely toxic, the general trend discovered nationwide was that they did not usually occur above trace readings. The scale ranged from a low of 17.0 to a high of 60.0. A preponderance of cases clustered around a score of 31.7 (the mean). The standard deviation was 8.93. As with the aesthetic/economic composite, households with higher scores generally had worse conditions—in this case, higher health risks—than households with lower scores.

BIVARIATE RELATIONSHIPS

Household attributes

None of the measures of household characteristics showed an association with the composite above an R^2 of 0.02 (Tables X-43 and X-44). As with the

Table X-43

Correlations Between the Health-Related Composite and Continuous Household Variables

Variable	r	R^2
INCOME	-0.13	0.01
PEOPLE	-0.01	0.00
EDUCATION	-0.05	0.00
AGE	0.01	0.00
RESIDENCE	-0.02	0.00
UNITS	-0.14	0.02

N = 21,974,000
n = 267

Table X-44

Summary Statistics for the Health-Related Composite According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.01
Full time	67.7	31.40	8.74	
Part time	6.0	32.29	4.64	
Keeping house	5.1	31.96	10.96	
Retired	19.8	32.48	9.83	
Unemployed	1.4	36.11	7.94	
ETHNICITY				0.00
White	89.7	31.65	9.04	
Nonwhite	10.3	32.23	7.80	
TENURE				0.01
Own	79.6	32.02	8.99	
Rent	17.5	30.33	8.21	
Other	2.9	31.29	10.31	
MOBILE HOME				0.00
Not mobile home	88.8	31.81	9.02	
Mobile home	11.2	30.85	8.11	
PERCEIVED COST				0.01
No cost	3.8	29.78	5.11	
Moderate	81.2	32.19	8.97	
Expensive	15.0	29.38	8.63	
WILLINGNESS TO PAY				0.01
Less	2.3	28.57	9.92	
Same	66.9	31.40	8.43	
More	30.8	32.60	9.76	

N = 21,974,000
n = 267

aesthetic/economic composite, the changes in average composite scores across variable categories tended to be minimal. The largest differences (four mean units) precluded comment since they were associated with categories containing very small proportions of the data. There did not appear to be a national, systematic association between any attributes of households and the distribution of scores on the health-related composite.

Supply attributes

In relating supply characteristics to the health composite, Tables X-45 and X-46 showed no coefficients of association sufficiently high to warrant regression modeling. Only supply type registered above zero, with an eta^2 of 0.02. Heretofore, the pattern of mean differences among supply types indicated that wells were slightly more likely to provide poorer quality water. In this instance, however, the small differences in the means pointed in the opposite direction. Community water supplies appeared to be slightly more likely to contain the trace element items of this composite. While the mean differences were opposite from the pattern registered by other measures of quality, they were not sufficiently great to warrant a conclusion. No households served by "other" types of supplies occurred in the NSA subsample.

System attributes

Tables X-47 and X-48 showed no indications of association sufficient to warrant further investigation.

INTERMEDIATE AND FINAL MODEL SPECIFICATION

The distribution of the trace toxic contaminants measured in the NSA health-related composite did not manifest a patterned association with the

Table X-45

Correlations Between the Health-Related Composite and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.03	0.00
RELIABILITY	0.04	0.00
ACCESSIBILITY	0.03	0.00

N = 21,974,000
n = 267

Table X-46

Summary Statistics for the Health-Related Composite According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.02
Community	53.9	32.64	8.23	
Well	46.1	30.26	9.59	

N = 21,974,000
n = 267

Table X-47

Correlations Between the Health-Related Composite and Continuous System Variables

Variable	r	R^2
CONNECTIONS	0.10	0.01
SERVICE YEARS	0.04	0.01

N = 21,974,000
n = 267

Table X-48

Summary Statistics for the Health-Related Composite According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.02
Private groups and individuals	49.5	30.60	9.39	
Public	43.5	32.58	8.65	
Private firm	7.0	34.16	8.65	
WATER TESTING				0.02
No testing	29.4	31.11	8.89	
Household tests	15.1	29.09	9.60	
System tests, or both household and system tests	55.6	32.73	8.53	

N = 21,974,000
n = 267

household, supply, or system characteristics. No intermediate nor final regression models were justified.

General Water Quality Composite

The third composite directly addressed the Congressional query about overall household water quality. This composite included all of the constituents measured. Standard scoring (Z scores) was used in the scale construction. Again, higher scores indicated generally poorer quality water. Since many of the constituents were measured on the 10 percent subsample, this general indicator was necessarily projected only from that subsample. The general water quality composite provided a general summary of household water conditions. It was a tool which allowed investigation into the broad question of the general state of rural water conditions. It was useful for suggesting general associations between overall water quality and household, supply, and system characteristics. The scale construction is specified in detail in Chapter VI.

BIVARIATE RELATIONSHIPS

Household attributes

Three household characteristics displayed associations which met or exceeded the R^2 or eta^2 criterion of 0.03: income, perceived cost, and willingness to pay more for an ideal water supply (see Tables X-49 and X-50). None of the associations was strong. Poorer people tended to have poorer water quality. Households which reported that their water supply cost nothing had higher average scores, representing lower quality water, than households which reported a cost for the water. In fact, the more costly the supply was perceived to be, the better was the average general quality of water. Similarly, households which were willing to pay more money if they could get an improved supply were households with the

Table X-49

Correlations Between the General Water Quality Composite and Continuous Household Variables

Variable	r	R^2
INCOME	-0.20	0.04*
PEOPLE	-0.01	0.00
EDUCATION	-0.15	0.02
AGE	0.12	0.02
RESIDENCE	0.10	0.01
UNITS	-0.08	0.01

N = 21,974,000
n = 267

*Underscoring denotes that the variable was retained for further analysis.

Table X-50

Summary Statistics for the General Water Quality Composite According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.01
Full time	67.7	8.64	6.90	
Part time	6.0	12.59	7.55	
Keeping house	5.1	9.32	6.42	
Retired	19.8	9.31	8.69	
Unemployed	1.4	8.42	2.75	
ETHNICITY				0.00
White	89.7	9.07	7.48	
Nonwhite	10.3	8.38	5.38	
TENURE				0.00
Own	79.6	9.05	7.52	
Rent	17.5	8.40	5.92	
Other	2.9	11.90	7.98	
MOBILE HOME				0.00
Not mobile home	88.8	9.01	7.23	
Mobile home	11.2	8.93	7.79	
PERCEIVED COST				0.03*
No cost	3.8	12.79	9.95	
Moderate	81.2	9.24	7.16	
Expensive	15.0	6.81	6.78	
WILLINGNESS TO PAY				0.04
Less	2.3	5.54	3.68	
Same	66.9	8.14	6.14	
More	30.8	11.11	9.11	

N = 21,974,000
n = 267

---

*Underscoring denotes that the variable was retained for further analysis.

poorest average water quality, while households which reported they would like to pay less registered the best average water quality among the three categories of willingness to pay.

Supply attributes

Among the supply characteristics examined, only the accessibility index met the criterion of 0.03, and therefore was retained for intermediate model specification. It appeared (Tables X-51 and X-52) that less accessible water supplies tended to provide poorer quality water.

System attributes

Tables X-53 and X-54 present those NSA variables which were measured for every system. There was no manifest association of sufficient strength to warrant inclusion in a regression model.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

Even though the associations between the individual variables and the general water quality composite were low, they were relatively independent. Table X-55 demonstrates that the three variables combined accounted for 10 percent of the variability in general water quality. The low coefficient for income indicated that income was redundant with the other variables.

Many of the constituents measured in .the NSA reach toxic levels long before they become detectable to the human senses. Others are more important for their aesthetically objectionable characteristics. Even for these latter contaminants, the laboratory measures are far more sensitive than the human senses. The general water quality scale comprised all constituents. It was somewhat surprising, then, that the perceptions of rural residents in terms of their

Table X-51

Correlations Between the General Water Quality Composite and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	-0.08	0.01
RELIABILITY	0.10	0.01
ACCESSIBILITY	0.17	0.03*

N = 21,974,000
n = 267

*Underscoring denotes that the variable was retained for further analysis.

Table X-52

Summary Statistics for the General Water Quality Composite According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.01
Community	53.9	8.30	6.06	
Well	46.1	9.75	8.34	

N = 21,974,000
n = 267

Table X-53

Correlations Between the General Water Quality Composite and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.14	0.02
SERVICE YEARS	0.09	0.01

N = 21,974,000
n = 267

Table X-54

Summary Statistics for the General Water Quality Composite According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.01
Private groups and individuals	49.5	9.51	8.31	
Public	43.5	8.17	6.04	
Private firm	7.0	10.55	6.07	
WATER TESTING				0.01
No testing	29.4	9.33	7.26	
Household tests	15.1	10.28	9.91	
System tests, or both household and system tests	55.6	8.47	6.36	

N = 21,974,000
n = 267

Table X-55

Intermediate Regression Estimates for the General Water Quality Composite and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	10.475
	WILLINGNESS TO PAY		
WP_1	Less	b_1	-2.600
WP_2	More	b_2	2.776
X_1	INCOME	b_3	-0.000
	PERCEIVED COST		
PC_1	No cost	b_4	2.144
PC_2	Moderate	b_5	0.084

$R^2 = .10$

Adjusted $R^2 = .10$

Standard error of estimate = 6.930

N = 21,974,000
n = 267

$$Y' = a + b_1WP_1 + b_2WP_2 + b_3X_1 + b_4PC_1 + b_5PC_2$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

willingness to pay and the perceived relative cost of their water supply contributed as much as they did to the association with water quality.

Intermediate supply model

As indicated earlier, among the supply variables considered, only accessibility met the 0.03 criterion of association with the general water quality composite. An intermediate regression model was not justified for one variable, since it would simply reiterate the relationship seen in Table X-51.

Intermediate system model

None of the system characteristics was associated with the general water quality composite. No intermediate model was necessary.

FINAL MODEL SPECIFICATION

The combination of household characteristics, supply accessibility, and the sample design factors (region, SMSA, and size of place) registered an adjusted R^2 of 0.25. Knowledge of these factors produced a moderate national pattern of association with the general quality of rural water supplies (Table X-56).

Among the design factors, region was the most meaningful nationwide correlate of general water quality. The coefficients suggest the same pattern discussed in the summary to the laboratory water quality presentation of Chapter V. The pattern of exceeded NSA reference values suggested that the Northeastern portion of the US usually had better quality water than the other regions. Households of the other regions which had less accessible supplies had an average water quality composite score of 13.5. The reference value described in Chapter VI was 9.0. Among households of this profile, 53.3 percent exceeded the reference value. Across the nation as a whole, 32.2 percent exceeded the reference value.

Table X-56

Final Regression Estimates for the General Water Quality Composite

Variable Designation		Parameter	Estimate
	Intercept	a	9.136
	REGION		
R_1	Northeast	b_1	-3.493
R_2	North Central	b_2	3.001
R_3	South	b_3	-2.321
	SMSA/NonSMSA		
S_1	SMSA	b_4	0.347
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	0.762
SI_2	Small rural communities	b_6	-0.872
	WILLINGNESS TO PAY		
WP_1	Less	b_7	-1.921
WP_2	More	b_8	2.414
X_1	INCOME	b_9	-0.000
X_2	ACCESSIBILITY	b_{10}	0.143
	PERCEIVED COST		
PC_1	No cost	b_{11}	0.977
PC_2	Moderate	b_{12}	0.458

$R^2 = 0.26$

Adjusted $R^2 = 0.25$

Standard error of estimate = 6.288

Table X-56 continued

N = 21,974,000
n = 267

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WP_1 + b_8WP_2 + b_9X_1 + b_{10}X_2 + b_{11}PC_1 + b_{12}PC_2$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

There were moderate associations between general water quality and geographical region, certain household attributes, and supply accessibility.

Perceived Water Quality

A truly comprehensive national profile of rural water quality must include the perspective of how rural residents evaluate their water quality. The absence of that perspective could increase the likelihood that socially inappropriate technologies might, at some point, be proposed.[6,7] Awareness of perceived water quality should help avert the "tyranny of expertise" in which knowledge of physical conditions becomes the sole basis for decisions regardless of the social setting.[8,9] Inattention to consumer preferences in the formation of drinking water standards has been an unfortunate historical pattern.[9]

The way that rural residents perceive their water supplies defines, in part, how they use it. W. I. Thomas stated that principle more generally when he said, "If men define situations as real, they are real in their consequences."[10] If rural residents perceive their water supply to be healthful, they will use it accordingly. They will not be reticent about drinking it, cooking with it, cleaning with it, or using it freely for any other purpose—even though the water may contain some harmful contaminant. On the other hand, should the residents consider the supply unhealthful or objectionable, they will avoid uses which they think could be detrimental—even though the water may be free of any danger.

The approach taken by the NSA for the questions used to construct the perceived water quality composite was to ask about the household's experience with particular attributes of the water: odor, taste, color, sediment, and cloudiness. People's evaluations tend to be consistent and stable when, as in the NSA, the referent is a "well-anchored physical property."[11]

Water chemistry laboratories were used by the NSA to develop indicator measures of water quality. The NSA also used household residents as sensing devices of water quality. Residents' evaluations are not as specific nor finely calibrated as laboratory equipment, but they have the advantage of historical experience—multiple readings—which were beyond the scope of the NSA's laboratory investigations.

The differences in sensitivity of laboratory measures and residents' evaluations were so great that agreement between them should be most apparent at the extremes. Very poor quality water should register poorly by both measurement methods. Table X-57 demonstrates the fundamental compatability of the two types of measurement. Each household was classified according to the size of the system serving it—individual (one connection), intermediate (two through fourteen connections), and community (fifteen or more connections). Each household was also classified as to the type of system residents would prefer to have. Water quality was only one of many possible reasons why a change in system might be desired, but it was one of the more important reasons. Over 80 percent of households served by each size of system preferred not to change. Those which did desire a change tended to have higher perceived quality scores, indicating a lower evaluation of the water, than those households whose residents were satisfied with their current system. Also, households preferring a change in system size had worse scores on the aesthetic/economic water quality composite. The same was generally the case for turbidity.

Table X-57 additionally shows the tendency for systems with fewer connections to be evaluated more positively. That finding concurs with surveys done by the Gallup organization which found that springs and wells were the most preferred supplies.[12]

A slight inversion of the association between the perceived quality and the aesthetic/economic composite resulted from grouping by system size category.

Table X-57

Average Scores on the Perceived Water Quality and Aesthetic/ Economic Composites, and Average Turbidity Concentrations for Households on Individual, Intermediate, or Community Systems According to Their Preferences for System Size

	Percent of Households	Average Perceived Water Quality Index	Average Aesthetic/ economic Composite	Average Turbidity
System preferences of households served by individual systems (8,765,000)				
Individual	88.2	2.52	2.45	2.26
Intermediate	1.8	6.29	3.14	3.43
Community	10.0	4.30	2.54	3.53
Total	100.0	2.79	2.48	2.41
System preferences of households served by intermediate systems (2,227,000)				
Individual	5.2	6.17	2.84	7.92
Intermediate	86.7	2.83	1.99	1.48
Community	8.1	4.61	2.89	2.14
Total	100.0	2.96	2.11	1.87
System preferences of households served by community systems (10,981,000)				
Individual	7.7	6.62	2.29	0.79
Intermediate	2.6	9.58	2.91	0.54
Community	89.7	3.34	1.80	0.60
Total	100.0	3.78	1.87	0.61

The overall association between the actual water quality (aesthetic/economic indicator) and the perceived quality was positive. That is, household perceptions of the quality of the water agreed with the composite score of laboratory-measured quality.

BIVARIATE RELATIONSHIPS

Laboratory-measured constituents

Among the water constituents measured in the NSA, only a few showed an association with rural residents' perceptions of water that were greater than the 0.03 criterion. As shown in Table X-58, six items had R^2s of 0.03 or greater, and two of them were composites. Conductance, iron, sulfate, and turbidity each were related to perceived water quality with an R^2 of 0.03. The aesthetic/economic indicator of water quality and the general water quality composite had R^2s of 0.06 and 0.05, respectively. In each case, the correlation was positive. As the contamination became greater, it was more likely that the household residents reported a problem. The association shown with specific conductance (used to estimate total dissolved solids) agreed with findings by William Bruvold that higher levels of total dissolved solids were associated with declining acceptability of the water to consumers.[14]

The R^2 of 0.05 between the general water quality composite and perceived water quality was largely attributable to those constituents which made up the aesthetic/economic composite. The latter composite was more parsimonious and the better measure to retain for modeling.

Household attributes

There were three variables among the household characteristics which showed some relationship to perceived water quality. Tables X-59 and X-60 show there was a slight tendency for younger heads of household and people who were in

Table X-58

Correlations Between Perceived Water Quality and Continuous Water Quality Constituents

Variable	r	R^2
Coliform	-0.01	0.00
Magnesium	0.10	0.01
Manganese	0.05	0.00
Color	0.11	0.01
Specific conductance	0.18	0.03
Iron	0.18	0.03
Lead	-0.01	0.00
Calcium	0.12	0.01
Nitrate-N	-0.06	0.00
Sodium	0.14	0.02
Sulfate	0.17	0.03
Turbidity	0.16	0.03*
Hardness	0.12	0.01
Aesthetic/economic composite	-0.24	-0.06
Arsenic	0.10	0.01
Barium	0.07	0.01
Cadmium	-0.01	0.00
Chromium	-0.05	0.00
Fluoride	0.03	0.00
Mercury	0.01	0.00
Selenium	0.03	0.00
Silver	-0.04	0.00
Gross alpha	0.03	0.00
Gross beta	-0.06	0.00
Methoxychlor	0.09	0.01
Lindane	0.01	0.00
Health-related composite	0.16	0.02
General water quality composite	0.21	0.05

*Underscoring denotes that the variable was retained for further analysis.

Table X-59

Correlations Between Perceived Water Quality Composite and Continuous Household Variables

Variable	r	R^2
INCOME	0.03	0.00
PEOPLE	0.06	0.00
EDUCATION	0.11	0.01
AGE	-0.19	0.04*
RESIDENCE	-0.17	0.03
UNITS	0.05	0.00

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table X-60

Summary Statistics for the Perceived Water Quality Composite According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.02
Full time	66.3	3.49	3.80	
Part time	4.8	4.05	4.17	
Keeping house	4.8	2.78	3.62	
Retired	21.3	2.52	3.36	
Unemployed	2.9	4.21	4.86	
ETHNICITY				0.00
White	91.4	3.22	3.73	
Nonwhite	8.6	4.14	4.30	
TENURE				0.01
Own	79.4	3.15	3.71	
Rent	17.8	4.06	4.10	
Other	2.7	3.17	3.09	
MOBILE HOME				0.00
Not mobile home	87.7	3.29	3.76	
Mobile home	12.3	3.41	4.01	
PERCEIVED COST				0.01
No cost	4.5	3.80	4.34	
Moderate	83.8	3.09	3.65	
Expensive	11.7	3.79	3.82	
WILLINGNESS TO PAY				0.07*
Less	3.6	3.03	3.51	
Same	63.4	2.59	3.21	
More	33.0	4.69	4.41	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

residence for a shorter time to give poorer evaluations of their supplies. The strongest association was with the apparent willingness of the respondent to pay for a better supply. Those who tended to report water problems were also more likely to say they were willing to pay more money if they could get a better supply. Saunders and Warford have argued that willingness to pay for an improved supply is a useful representation of the value placed on the benefits of a better supply.[15] NSA findings suggest that willingness to pay also reflected the physical quality of the water in terms of the aesthetic/economic constituents.

Supply attributes

None of the supply attributes measured for all supplies displayed a notable association with perceived water quality. Tables X-61 and X-62 show that none of the squared measures of association was above 0.02.

System attributes

As with supply characteristics, none of the attributes of rural water systems showed an R^2 or an eta^2 higher than 0.02 (Tables X-63 and X-64).

INTERMEDIATE MODEL SPECIFICATION

Intermediate constituent model

The aesthetic/economic composite and the measure of turbidity were retained for inclusion in the intermediate model. Other constituents met the minimum criterion of 0.03, but they were redundant with the aesthetic/economic composite.

Table X-65 indicates that these water quality parameters registered a combined R^2 of 0.06 with the perceived water quality composite.

Table X-61

Correlations Between the Perceived Water Quality Composite and Continuous Supply Variables

Variable	r	R^2
TOTAL TREATMENT	0.06	0.00
RELIABILITY	0.15	0.02
ACCESSIBILITY	-0.04	0.00

N = 21,974,000
n = 2,654

Table X-62

Summary Statistics for the Perceived Water Quality Composite According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.01
Community	50.0	3.76	3.91	
Well	44.8	2.82	3.64	
Other	5.2	1.37	1.80	

N = 21,974,000
n = 2,654

Table X-63

Correlations Between the Perceived Water Quality Composite and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.02	0.00
SERVICE YEARS	0.12	0.01

N = 21,974,000
n = 2,654

Table X-64

Summary Statistics for the Perceived Water Quality Composite According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.01
Private groups and individuals	52.3	2.91	3.67	
Public	40.7	3.69	3.86	
Private firm	6.7	4.08	3.93	
Other	0.3	1.37	1.80	
WATER TESTING				0.02
No testing	29.8	92.76	3.44	
Household tests	17.3	2.85	3.77	
System tests, or both household and system tests	52.9	3.76	3.92	

N = 21,974,000
n = 2,654

Table X-65

Intermediate Regression Estimates for the Perceived Water Quality Composite and Laboratory-Measured Constituents

Variable Designation		Parameter	Estimate
	Intercept	a	2.647
X_1	AESTHETIC/ECONOMIC COMPOSITE	b_1	0.274
X_2	TURBIDITY	b_2	0.046

$R^2 = 0.06$

Adjusted $R^2 = 0.06$

Standard error of estimate = 3.667

N = 21,974,000
n = 2,654

$Y' = a + b_1X_1 + b_2X_2$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Intermediate household model

The three household variables (willingness to pay, age of head, and length of residence) together showed an R^2 of 0.09 with the perceived water quality composite (see Table X-66). As was also shown with the bivariate associations, willingness to pay was the dominant variable.

Intermediate supply model

None of the supply characteristics showed a relationship of sufficient strength to require modeling.

Intermediate system model

Perceived water quality was not related to any of the system variables, so an intermediate system model could not be developed.

FINAL MODEL SPECIFICATION

The findings regarding supply and system characteristics indicated that perceived water quality tended not to be related to the physical extraction, conveyance, or transportation apparatus of the supply, nor to aspects of the organizational system.

The combination of sample design variables, water quality measures, and household characteristics shown in Table X-67 accounted for 16 percent of the variance of perceived water quality. Among the design characteristics, two categories showed some importance. Large rural communities tended to have higher perceived quality. Households in the North Central tended to have low perceived quality. Among those variables which measure physical properties, the strongest association with rural residents' perceptions of water quality was the composite indicator of laboratory-measured aesthetic and economic components.

Table X-66

Intermediate Regression Estimates for the Perceived Water Quality Composite and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	4.845
	WILLINGNESS TO PAY		
WP_1	Less	b_1	-0.319
WP_2	More	b_2	1.908
X_1	AGE	b_3	-0.025
X_2	RESIDENCE	b_4	-0.023

$R^2 = 0.09$

Adjusted $R^2 = 0.09$

Standard error of estimate = 3.605

N = 21,974,000
n = 2,654

$$Y' = a + b_1WP_1 + b_2WP_2 + b_3X_1 + b_4X_2$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table X-67

Final Regression Estimates for the Perceived Water Quality Composite

Variable	Designation	Parameter	Estimate
	Intercept	a	5.044
	REGION		
R_1	Northeast	b_1	0.042
R_2	North Central	b_2	-0.334
R_3	South	b_3	-0.006
	SMSA/NonSMSA		
S_1	SMSA	b_4	0.004
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	0.701
SI_2	Small rural communities	b_6	-0.022
	WILLINGNESS TO PAY		
WP_1	Less	b_7	-0.193
WP_2	More	b_8	0.912
X_1	AESTHETIC/ECONOMIC COMPOSITE	b_9	0.254
X_2	AGE	b_{10}	-0.031
X_3	RESIDENCE	b_{11}	-0.022
X_4	TURBIDITY	b_{12}	0.061

$R^2 = 0.17$

Adjusted $R^2 = 0.16$

Standard error of estimate = 3.459

Table X-67 continued

N = 21,974,000
n = 2,654

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WP_1 + b_8WP_2 + b_9X_1 + b_{10}X_2 + b_{11}X_3 + b_{12}X_4$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

SUMMARY AND CONCLUSIONS

Seven indicators of water quality, based on laboratory findings, have been described: total coliform concentration, nitrate-N, lead, turbidity, an aesthetic/economic index, a health-related index, and a general water quality indicator. In addition, one indicator of perceived water quality has been discussed. For each of these indicators of water quality, household, supply, and system characteristics have been examined in search of national patterns that would indicate systematic relationships.

As to total coliform, the final regression model indicated that the type of supply was the most important correlate of coliform concentrations. Community supplies tended to have lower concentrations than wells or other miscellaneous types of supply. But, considering all of the relevant variables, the proportion of the variation of total coliform concentrations which was accounted for was only 19 percent. This implied that the household, supply, and system variables were only moderately useful for hypothesizing which water supplies were likely to be contaminated. They did help focus attention on certain configurations that had higher contamination potential, however. One such configuration included households where the head had a high school education or less, where the household income was under $15,000, where there was service by a system with fewer than 500 connections, where there were multiple treatment devices hooked to the supply, where supplies were less accessible than most, and where a supply had been in service for more than eight years. Almost 48 percent of households in this group had more than one organism per 100 milliliters of water, compared to about 29 percent nationally.

Less can be said about the correlates of nitrate-N concentrations. Only 7 percent of the variability was accounted for by knowing the household, supply, and system characteristics. Regional geography, as reported in Chapter V, was the most influential variable.

Lead appears to be fairly pervasive at low concentrations. Like nitrate-N, it appears to be independent of—and equally likely to affect—any type of household, supply, or system.

A total of only 6 percent of the turbidity variance was accounted for. Turbidity is a common problem affecting households, supplies, and systems of all types. Community supplies tended toward lower turbidity levels. Concentrations of turbidity in wells varied from very low to very high.

As to the aesthetic/economic indicator, there did not appear to be any systematic national pattern. An examination of means suggested that households which were willing to pay more for a better supply did tend to have poorer scores on the aesthetic/economic composite. Households served by wells tended to give poorer evaluations than those on any other type of supply.

Again, no systematic national pattern emerged for the health-related composite. The trace constituents represented by the indicator occurred for reasons that were apparently unrelated to the household, supply, and system characteristics examined.

Of all the indicators of water quality, the general water quality indicator was the most useful for classifying households. The combined model showed that the sample design characteristics, a household's willingness to pay more for a better supply, household income, supply accessibility, the number of water-using devices, and the perceived relative cost of the water could account for 25 percent of the variability of the general water quality indicator. The two most important characteristics were the geographical region (the Northeast being generally better off) and the household's willingness to pay more for a better supply.

In regard to the perceived water quality indicator, rural residents tended to display strong consistency in their responses to a variety of questions about their water conditions. Also, households which assessed the quality of their supplies as poor tended to have water which laboratory measurements indicated as more likely to have aesthetic/economic problems. Households' willingness to pay for a better supply was another useful indicator of their overall perception of water quality. Those willing to pay more tended to give poorer evaluations of the supply, which were also reflected in scores on the aesthetic/economic indicator.

The development of closed water systems, disinfection, and source protection have had a fairly strong stabilizing effect in improving the general quality of domestic US water supplies. Since the majority of systems in rural America (individual wells) had conventional pumping and piping arrangements with little or no treatment, it follows that the character of the water in the source was probably the predominant influence upon the quality of the delivered water.

The average rural US household has relatively good water, as indicated by NSA parameters. Poor water quality, while not rare, does not have a strong association at the national level with any household, supply, or system characteristic studied here.

WATER QUALITY AMONG RURAL HOUSEHOLDS SERVED BY WELLS: A REGRESSION APPROACH

Almost 45 percent of all rural households were served by wells which had fewer than fifteen connections. They are singled out for additional attention because of configurations among well supplies and the lack of information regarding the possible effects of those configurations upon the quality of the water.

The following analysis parallels the preceding presentation by examining attributes of households, supplies, and systems as they were associated with various measures of water quality, with the omission of perceived quality. Since the analysis deals with wells, several supply variables were added (Table X-68) to distinguish various well configurations.

Table X-68

Additional Independent Variables Used in the Analysis of Water Quality (For Households Served by Wells)

Continuous Variables	Categorical Variables
Supply Attributes	
WELL DEPTH	WELL TYPE (dug, bored, drilled, driven, jetted, other)
WELL AGE	
DISTANCE TO CONTAMINATION	GROUTING (grouted, not grouted)
	WELL EXIT (pitless adapter, buried seal, terminates above ground, unclassified)
	SPLASHPROOF (yes, no, unclassified)
	COVER (inadequate, adequate, unclassified)
	MAINTENANCE (inadequate, adequate)

Total Coliform

BIVARIATE RELATIONSHIPS

Household attributes

Table X-69 shows the correlations between the counts of total coliform bacteria and continuous measures of household attributes. Only two squared associations are greater than zero, income and education. Neither is sufficiently large to warrant further investigation. Table X-70 displays a similar pattern among categorical household variables. None is associated with coliform contamination sufficiently to be retained for regression analysis. None of the household variables considered showed an R^2 or eta^2 association with total coliform counts of sufficient magnitude to justify inclusion in a regression model.

Supply attributes

All of the continuous supply variables demonstrated associations with total coliform below an R^2 of 0.03 (see Table X-71). Four categorical supply characteristics had eta^2 values at 0.03 or higher. The place from which the water exited the well did show some association ($eta^2 = 0.03$) with total coliform counts (see Table X-72). Wells using pitless adapters tended to have lower average coliform counts than those which had buried seals. Both wells with pitless adapters and those with buried seals had lower counts than those which were designed to have the water exit the well casing above ground.

Wells with covers judged to be splashproof tended to have lower average total coliform counts than those which were not. The association between the classification of the cover and the coliform count was an eta^2 of 0.04.

Whether or not the well was grouted had an association with total coliform counts at an eta^2 of 0.12. Grouted wells tended to be associated with lower counts.

X-69

Correlations Between Total Coliform Counts and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	-0.11	0.01
PEOPLE	0.01	0.00
EDUCATION	-0.13	0.02
AGE	-0.00	0.00
RESIDENCE	0.04	0.00
UNITS	0.05	0.00

N = 9,850,000
n = 1,315

Table X-70

Summary Statistics for Total Coliform According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.7	10,507	100,732	
Part time	4.4	29,210	168,023	
Keeping house	5.9	17,271	129,890	
Retired	20.1	3,496	56,839	
Unemployed	3.0	20,022	137,847	
ETHNICITY				0.00
White	93.0	8,806	92,117	
Nonwhite	7.0	34,404	181,426	
TENURE				0.00
Own	80.6	11,199	104,440	
Rent	15.4	6,272	74,293	
Other	4.0	15,379	122,746	
MOBILE HOME				0.00
Not mobile home	87.8	10,166	99,042	
Mobile home	12.2	13,787	115,728	
PERCEIVED COST				0.00
No cost	4.9	56,667	230,746	
Moderate	88.5	6,959	81,520	
Expensive	5.7	29,094	167,962	
WILLINGNESS TO PAY				0.01
Less	3.5	251	952	
Same	66.6	10,189	99,831	
More	29.9	12,767	109,863	

N = 9,850,000
n = 1,315

Table X-71

Correlations Between Total Coliform and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	-0.09	0.01
RELIABILITY	0.07	0.00
ACCESSIBILITY	0.15	0.02
SUPPLY DISTANCE	0.00	0.00
WELL DEPTH	-0.14	0.02
DISTANCE TO CONTAMINATION	-0.06	0.00
WELL AGE	0.04	0.00

N = 9,850,000
n = 1,315

Table X-72
Summary Statistics for Total Coliform According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.15*
Dug	12.7	31,154	170,211	
Bored	6.2	318	1,271	
Drilled	67.1	8,977	93,776	
Driven	11.7	5,267	72,120	
Jetted	1.3	80	213	
Other	1.0	12	36	
WELL COVER				0.01
Adequate	77.1	9,397	95,857	
Inadequate	22.9	14,683	117,438	
WELL EXIT				0.03
Top of ground	51.5	14,062	116,158	
Pitless adapter	28.9	6,103	77,003	
Buried Seal	18.3	8,733	92,766	
Unclassified	1.4	10	32	
SPLASHPROOF				0.04
No	21.1	15,627	122,121	
Yes	78.9	9,262	94,809	
GROUTING				0.12
Grouted	42.5	6,750	81,235	
Not grouted	57.5	13,461	113,696	
MAINTENANCE				0.02
Well maintained	78.7	9,243	94,727	
Poorly maintained	21.3	15,644	122,143	
DISPOSAL SYSTEM				0.01
Septic	90.7	8,516	90,732	
Public sewer	3.4	32,017	175,939	
Other	5.9	30,608	169,530	

N = 9,850,000
n = 1,315

*Underscoring denotes that the variable was retained for further analysis.

The strongest association, eta^2 of 0.15, was shown between total coliform counts and the type of well construction. Some of the construction modes were so rare that interpretation of their average coliform counts was difficult, but dug wells clearly were the most susceptible to high total coliform counts. Some construction techniques, such as boring, jetting, and driving, are restricted from wholesale application because they generally require special soil and geologic conditions. Sandy, stone-free conditions are favorable for these techniques. These same conditions enhance the filtering and, therefore, the protection potential for groundwater against such contaminants as coliform bacteria. This site, self selectiveness of the construction technique may be related to the generally lower average counts demonstrated by jetted, bored, and driven wells.

System attributes

None of the measured system characteristics of Table X-73 or X-74 demonstrated associations with total coliform counts sufficient to warrant discussion.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

None of the household variables were retained for regression analysis.

Intermediate supply model

Four variables, well type, grouting, splashproof, and well exit, were included in the supply intermediate regression model. For purposes of the analysis, the group of wells that were "unclassified" were grouped with the category "the water exits the casing above ground." Table X-75 shows that their combined effect was only slightly greater than that demonstrated by the well type alone. The four variables in concert accounted for an R^2 of 0.16 with total coliform, whereas well

Table X-73

Correlations Between Total Coliform and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	-0.04	0.00
YEARS SERVED	0.11	0.01

N = 9,850,000
n = 1,315

Table X-74

Summary Statistics for Total Coliform According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	10,921	102,699	
Public	2.0	10	34	
Private firm	0.4	1	3	
Other	0.5	16	29	
WATER TESTING				0.00
No testing	57.0	13,840	115,470	
Household tests	36.5	7,427	89,904	
System tests, or both household and system tests	6.5	125	740	

N = 9,850,000
n = 1,315

Table X-75

Intermediate Regression Estimates for Total Coliform and Supply Attributes (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	2.021
	WELL TYPE		
WT_1	Dug	b_1	2.075
WT_2	Bored	b_2	0.048
WT_3	Drilled	b_3	-0.632
WT_4	Driven	b_4	-0.966
WT_5	Jetted	b_5	0.114
G_1	GROUTING	b_6	-0.080
SP_1	SPLASHPROOF	b_7	-0.147
	WELL EXIT		
WE_1	Top of ground	b_8	0.339
WE_2	Pit	b_9	-0.236

$R^2 = 0.16$

Adjusted $R^2 = 0.16$

Standard error of estimate = 2.40

N = 9,850,000
n = 1,315

$$Y' = a + b_1WT_1 + b_2WT_2 + b_3WT_3 + b_4WT_4 + b_5WT_5 + b_6G_1 + b_7SP_1 + b_8WE_1 + b_9WE_2$$

type alone had an eta^2 of 0.15. The demonstrated overlap is consistent with the usual expectations for some well types. Dug wells, for example, tend not to be grouted, they are not usually immune to surface water intrusion (not splashproof) and the water exits the well above ground—all conditions associated with high average total coliform counts.

Intermediate system model

None of the system attributes were retained for regression analysis with total coliform counts.

FINAL MODEL SPECIFICATION

Table X-76 combines the four supply attributes which had sufficient association with total coliform counts with the design variables in a final total coliform model for households served by wells. Table X-76 shows that much like the intermediate supply model, the addition of other variables did not increase the association much beyond that demonstrated between well type and total coliform counts alone. The well type eta^2 was 0.15. The final model had an R^2 of 0.17.

The R^2 of 0.17 indicates that the range of coliform counts encountered in rural wells was not strongly associated with the measured household, supply, or system attributes. The well type was the single best indicator among those examined.

Turbidity

BIVARIATE RELATIONSHIPS

Household attributes

The association between most of the measured household characteristics and the level of turbidity in well water was usually zero. Tables X-77 and X-78

Table X-76

Final Regression Estimates for Total Coliform
(For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	1.823
	REGION		
R_1	Northeast	b_1	-0.196
R_2	North Central	b_2	-0.303
R_3	South	b_3	0.413
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.280
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.259
SI_2	Small rural communities	b_6	-0.041
	WELL TYPE		
WT_1	Dug	b_7	2.164
WT_2	Bored	b_8	-0.069
WT_3	Drilled	b_9	-0.552
WT_4	Driven	b_{10}	-0.771
WT_5	Jetted	b_{11}	-0.022
G_1	GROUTING	b_{12}	-0.078
SP_1	SPLASHPROOF	b_{13}	-0.123
	WELL EXIT		
WE_1	Top of ground	b_{14}	0.214
WE_2	Pit	b_{15}	-0.139

$R^2 = 0.18$

Adjusted $R^2 = 0.17$

Standard error of estimate = 2.37

Table X-76 continued

N = 9,850,000

n = 1,315

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WT_1 + b_8WT_2 + b_9WT_3 + b_{10}WT_4 + b_{11}WT_5 + b_{12}G_1 + b_{13}SP_1 + b_{14}WE_1 + b_{15}WE_2$$

Table X-77

Correlations Between Turbidity and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	-0.10	0.01
PEOPLE	-0.05	0.00
EDUCATION	-0.14	0.02
AGE	0.09	0.01
RESIDENCE	0.07	0.01
UNITS	-0.03	0.00

N = 9,850,000
n = 1,315

Table X-78

Summary Statistics for Turbidity According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.01
Full time	66.7	2.05	6.77	
Part time	4.4	1.83	6.82	
Keeping house	5.9	2.73	7.01	
Retired	20.1	3.85	12.72	
Unemployed	3.0	2.73	10.37	
ETHNICITY				0.01
White	93.0	2.30	8.03	
Nonwhite	7.0	4.57	12.80	
TENURE				0.00
Own	80.6	2.58	8.56	
Rent	15.4	1.62	7.21	
Other	4.0	3.37	10.70	
MOBILE HOME				0.00
Not mobile home	87.8	2.38	8.36	
Mobile home	12.2	3.03	9.21	
PERCEIVED COST				0.00
No cost	4.9	3.36	10.67	
Moderate	88.5	2.43	8.37	
Expensive	5.7	2.03	8.32	
WILLINGNESS TO PAY				0.00
Less	3.5	2.06	5.45	
Same	66.6	2.15	7.87	
More	29.9	3.21	9.89	

N = 9,850,000
n = 1,315

indicate the highest association was an R^2 of 0.02 between turbidity and the level of education of the household head.

Supply attributes

Tables X-79 and X-80 are similarly void of association. None of the measured supply attributes was notably related to the levels of turbidity found among rural households served by wells.

System attributes

Four system attributes are displayed in Tables X-81 and X-82. None showed an association with turbidity greater than 0.00.

INTERMEDIATE AND FINAL MODEL SPECIFICATION

No household, supply, or system characteristic showed an association with levels of turbidity among rural households served by wells to justify investigation with regression models.

Nitrate-N

BIVARIATE RELATIONSHIPS

Household attributes

Continuous measures of household characteristics in Table X-83 and categorical measures in Table X-84 do not show any squared measures of association with nitrate-N larger than 0.01.

Table X-79

Correlations Between Turbidity and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	-0.06	0.00
RELIABILITY	0.06	0.00
ACCESSIBILITY	0.11	0.01
SUPPLY DISTANCE	-0.03	0.00
WELL DEPTH	-0.03	0.00
DISTANCE TO CONTAMINATION	0.10	0.01
WELL AGE	-0.02	0.00

N = 9,850,000
n = 1,315

Table X-80

Summary Statistics for Turbidity According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.00
Dug	12.7	1.70	5.52	
Bored	6.2	2.70	11.43	
Drilled	67.1	2.56	8.87	
Driven	11.7	2.54	7.20	
Jetted	1.3	3.47	6.53	
Other	1.0	1.83	5.06	
WELL COVER				0.00
Adequate	77.1	2.26	8.33	
Inadequate	22.9	3.12	8.89	
WELL EXIT				0.00
Top of ground	51.5	2.62	8.74	
Pitless adapter	28.9	2.20	8.80	
Buried Seal	18.3	2.46	7.28	
Unclassified	1.4	1.96	4.61	
SPLASHPROOF				0.01
No	21.1	3.43	10.11	
Yes	78.9	2.21	7.96	
GROUTING				0.00
Grouted	42.5	2.77	9.90	
Not grouted	57.5	2.23	7.22	
MAINTENANCE				0.01
Well maintained	78.7	2.11	7.78	
Poorly maintained	21.3	3.75	10.55	
DISPOSAL SYSTEM				0.01
Septic	90.7	2.31	8.21	
Public sewer	3.4	2.77	7.65	
Other	5.9	4.58	11.89	

N = 9,850,000
n = 1,315

Table X-81

Correlations Between Turbidity and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	0.02	0.00
YEARS SERVED	0.02	0.00

N = 9,850,000
n = 1,315

Table X-82

Summary Statistics for Turbidity According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	2.51	8.59	
Public	2.0	0.80	0.66	
Private firm	0.4	0.44	0.48	
Other	0.5	0.41	0.44	
WATER TESTING				0.00
No testing	57.0	2.59	9.00	
Household tests	36.0	2.45	8.11	
System tests, or both household and system tests	6.5	1.36	4.83	

N = 9,850,000
n = 1,315

Table X-83

Correlations Between Nitrate-N and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	0.03	0.00
PEOPLE	0.01	0.00
EDUCATION	-0.06	0.00
AGE	0.02	0.00
RESIDENCE	0.07	0.00
UNITS	-0.02	0.00

N = 9,850,000
n = 1,315

Table X-84

Summary Statistics for Nitrate-N
According to Categorical Household Variables
(For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.01
Full time	66.7	2.37	6.08	
Part time	4.4	3.87	8.52	
Keeping house	5.9	2.20	3.88	
Retired	20.1	2.36	6.37	
Unemployed	3.0	3.72	4.88	
ETHNICITY				0.00
White	93.0	2.49	6.24	
Nonwhite	7.0	2.12	4.52	
TENURE				0.00
Own	80.6	2.37	6.16	
Rent	15.4	2.72	4.51	
Other	4.0	3.42	9.87	
MOBILE HOME				0.00
Not mobile home	87.8	2.54	6.38	
Mobile home	12.2	1.98	4.00	
PERCEIVED COST				0.00
No cost	4.9	3.35	8.23	
Moderate	88.5	2.43	6.16	
Expensive	5.7	2.38	3.52	
WILLINGNESS TO PAY				0.01
Less	3.5	2.54	4.99	
Same	66.6	2.64	6.35	
More	29.9	2.06	5.75	

N = 9,850,000
n = 1,315

Supply attributes

None of the continuous supply variables of Table X-85 had an R^2 over 0.01 with nitrate-N. Among the categorical variables of Table X-86, only well type registered an eta^2 of 0.03. None were higher. Average nitrate-N levels tended to be greatest in dug and driven wells.

System attributes

Only years served had a squared association of 0.01. The other three characteristics of Tables X-87 and X-88 were 0.00.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

None of the variables was sufficiently associated with nitrate-N to justify a regression model.

Intermediate supply model

Since only well type met the minimum 0.03 squared association among the supply variables, a regression model with only well type as an independent variable is not necessary since it would replicate exactly the association already seen in Table X-86.

Intermediate system model

None of the supply variables was retained for further analysis.

FINAL MODEL SPECIFICATION

The supply attribute, well type, and the sample design variables demonstrated a combined association with nitrate-N of an adjusted R^2 of 0.05 (see Table X-89).

Table X-85

Correlations Between Nitrate-N and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	-0.05	0.00
RELIABILITY	0.00	0.00
ACCESSIBILITY	0.07	0.00
SUPPLY DISTANCE	-0.01	0.00
WELL DEPTH	-0.12	0.01
DISTANCE TO CONTAMINATION	-0.01	0.00
WELL AGE	0.02	0.00

N = 9,850,000
n = 1,315

Table X-86

Summary Statistics for Nitrate-N According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.03*
Dug	12.7	3.79	7.08	
Bored	6.2	2.09	3.14	
Drilled	67.1	2.12	5.91	
Driven	11.7	3.47	7.62	
Jetted	1.3	1.05	1.35	
Other	1.0	1.55	1.56	
WELL COVER				0.00
Adequate	77.1	2.31	6.00	
Inadequate	22.9	3.00	6.56	
WELL EXIT				0.00
Top of ground	51.5	2.52	5.65	
Pitless adapter	28.9	2.64	7.48	
Buried seal	18.3	1.94	4.75	
Unclassified	1.4	3.71	8.06	
SPLASHPROOF				0.00
No	21.1	2.93	6.60	
Yes	78.9	2.34	6.00	
GROUTING				0.02
Grouted	42.5	2.39	7.29	
Not grouted	57.5	2.52	5.12	
MAINTENANCE				0.00
Well maintained	78.7	2.47	6.30	
Poorly maintained	21.3	2.44	5.51	
DISPOSAL SYSTEM				0.00
Septic	90.7	2.41	6.19	
Public sewer	3.4	2.64	3.34	
Other	5.9	3.22	6.50	

N = 9,850,000
n = 1,315

*Underscoring denotes that the variable was retained for further analysis.

Table X-87

Correlations Between Nitrate-N and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	-0.03	0.00
YEARS SERVED	0.11	0.01

N = 9,850,000
n = 1,315

Table X-88

Summary Statistics for Nitrate-N According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	2.51	6.22	
Public	2.0	1.10	1.91	
Private firm	0.4	1.11	1.39	
Other	0.5	1.07	1.54	
WATER TESTING				0.00
No testing	57.0	2.47	5.48	
Household tests	36.0	2.65	7.43	
System tests, or both household and system tests	6.5	1.37	2.27	

N = 9,850,000
n = 1,315

Table X-89

Final Regression Estimates for Nitrate-N (For Households Served by Wells)

Variable	Designation	Parameter	Estimate
	Intercept	a	1.265
	REGION		
R_1	Northeast	b_1	-0.038
R_2	North Central	b_2	0.055
R_3	South	b_3	-0.127
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.063
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.041
SI_2	Small rural communities	b_6	-0.362
	WELL TYPE		
WT_1	Dug	b_7	0.271
WT_2	Bored	b_8	0.076
WT_3	Drilled	b_9	-0.113
WT_4	Driven	b_{10}	0.088
WT_5	Jetted	b_{11}	-0.110

$R^2 = 0.05$

Adjusted $R^2 = 0.05$

Standard error of estimate = 0.75

N = 9,850,000
n = 1,315

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WT_1 + b_8WT_2 + b_9WT_3 + b_{10}WT_4 + b_{11}WT_5$$

The generally weak associations indicate that the incidence of nitrate-N contamination among households served by wells was not systematically, linearly associated with either one or combinations of measured household, supply, and system characteristics.

Lead

BIVARIATE RELATIONSHIPS

Household attributes

Tables X-90 and X-91 show continuous and categorical household characteristics. None of them had a squared measure of association greater than 0.01.

Supply attributes

Among the supply characteristics of Tables X-92 and X-93, only one, well exit, had an association above the minimum 0.03 criterion value. Well exit had an eta^2 of 0.04 with measured lead levels. Wells in which the water exited the well above ground level had slightly higher lead levels, on the average, than those using pitless adapters and those with buried seals.

System attributes

No measured system characteristic (Tables X-94 and X-95) had squared measures of association above 0.00 with lead levels.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

No household attribute showed a systematic association with lead levels among households with well water supplies. No intermediate regression model was justified.

Table X-90

Correlations Between Lead and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	-0.06	0.00
PEOPLE	-0.02	0.00
EDUCATION	-0.07	0.00
AGE	-0.02	0.00
RESIDENCE	0.04	0.00
UNITS	0.01	0.00

N = 9,850,000
n = 1,315

Table X-91

Summary Statistics for Lead According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	66.7	0.03	0.04	
Part time	4.4	0.03	0.04	
Keeping house	5.9	0.02	0.03	
Retired	20.1	0.03	0.05	
Unemployed	3.0	0.04	0.05	
ETHNICITY				0.01
White	93.0	0.03	0.04	
Nonwhite	7.0	0.04	0.04	
TENURE				0.00
Own	80.6	0.03	0.04	
Rent	15.4	0.03	0.05	
Other	4.0	0.02	0.03	
MOBILE HOME				0.00
Not mobile home	87.8	0.03	0.05	
Mobile home	12.2	0.02	0.03	
PERCEIVED COST				0.01
No cost	4.9	0.04	0.06	
Moderate	88.5	0.03	0.04	
Expensive	5.7	0.04	0.08	
WILLINGNESS TO PAY				0.01
Less	3.5	0.02	0.04	
Same	66.6	0.03	0.05	
More	29.9	0.03	0.04	

N = 9,850,000
n = 1,315

Table X-92

Correlations Between Lead and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	0.00	0.00
RELIABILITY	0.03	0.00
ACCESSIBILITY	0.08	0.01
SUPPLY DISTANCE	0.05	0.00
WELL DEPTH	0.05	0.00
DISTANCE TO CONTAMINATION	0.00	0.00
WELL AGE	0.03	0.00

N = 9,850,000
n = 1,315

Table X-93

Summary Statistics for Lead According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.02
Dug	12.7	0.04	0.05	
Bored	6.2	0.04	0.05	
Drilled	67.1	0.02	0.04	
Driven	11.7	0.02	0.05	
Jetted	1.3	0.04	0.06	
Other	1.0	0.03	0.03	
WELL COVER				0.00
Adequate	77.1	0.03	0.04	
Inadequate	22.9	0.03	0.06	
WELL EXIT				0.04*
Top of ground	51.5	0.03	0.05	
Pitless adapter	28.9	0.02	0.04	
Buried Seal	18.3	0.02	0.04	
Unclassified	1.4	0.02	0.03	
SPLASHPROOF				0.01
No	21.1	0.03	0.06	
Yes	78.9	0.03	0.04	
GROUTING				0.01
Grouted	42.5	0.02	0.04	
Not grouted	57.5	0.03	0.05	
MAINTENANCE				0.00
Well maintained	78.7	0.03	0.04	
Poorly maintained	21.3	0.03	0.06	
DISPOSAL SYSTEM				0.01
Septic	90.7	0.03	0.05	
Public sewer	3.4	0.02	0.03	
Other	5.9	0.04	0.05	

N = 9,850,000
n = 1,315

*Underscoring denotes that the variable was retained for further analysis.

Table X-94

Correlations Between Lead and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	0.03	0.00
YEARS SERVED	0.05	0.00

N = 9,850,000
n = 1,315

Table X-95

Summary Statistics for Lead According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	0.03	0.05	
Public	2.0	0.03	0.03	
Private firm	0.4	0.02	0.03	
Other	0.5	0.01	0.01	
WATER TESTING				0.00
No testing	57.0	0.03	0.04	
Household tests	36.5	0.03	0.05	
System tests, or both household and system tests	6.5	0.03	0.05	

N = 9,850,000
n = 1,315

Intermediate supply model

Only one supply characteristic, well exit, surpassed the minimum squared association of 0.03. A one-variable model would be redundant with the information of Table X-93, and is therefore not presented.

Intermediate system model

None of the system variables was retained for regression analysis.

FINAL MODEL SPECIFICATION

The combination of well exit and the sample design variables accounts for 7.0 percent of the variability in lead levels (Table X-96). The indication was that lead levels in water supplies of households served by wells was not systematically, linearly associated with measured attributes of the households, supplies, or systems.

COMPOSITE INDICATORS OF WATER QUALITY

Two of the composite indicators, the health-related and the general water ality composite, were calculated for the 10 percent NSA subsample of house-olds. Since this analysis deals only with households served by wells, those composites represent a 10 percent subsample of the well subset. As a result, some of the percentages of cases in categories are not the same as those presented in earlier tables of this section.

Table X-96

Final Regression Estimates for Lead (For Households Served by Wells)

Variable	Designation	Parameter	Estimate
	Intercept	a	-4.911
	REGION		
R_1	Northeast	b_1	-0.322
R_2	North Central	b_2	-0.103
R_3	South	b_3	0.151
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.106
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.016
SI_2	Small rural communities	b_6	0.547
	WELL EXIT		
WE_1	Top of ground	b_7	0.223
WE_2	Pit	b_8	-0.11

$R^2 = 0.08$

Adjusted $R^2 = 0.07$

Standard error of estimate = 1.06

N = 9,850,000
n = 1,315

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WE_1 + b_8WE_2$$

Aesthetic/Economic Composite

BIVARIATE RELATIONSHIPS

Household attributes

No household variable measured had a strong association with the levels of the aesthetic/economic composite. Of the variables shown in Tables X-97 and X-98, the highest association was only an eta^2 of 0.02.

Supply attributes

Similarly, none of the supply variables of Tables X-99 and X-100 were found to be systematically associated with the aesthetic/economic composite.

System attributes

All four of the system characteristics shown in Tables X-101 and X-102 had squared associations of 0.00 with the aesthetic/economic composite.

INTERMEDIATE AND FINAL MODEL SPECIFICATION

No regression models were appropriate. There was no indication of any systematic national association between the aesthetic/economic indicator of water quality and any household, supply, or system characteristics.

Health-Related Composite

BIVARIATE RELATIONSHIPS

Household attributes

There was a slight association ($R^2 = 0.03$) between the education level of the head of household and the health-related water quality indicator (see Table

Table X-97

Correlations Between the Aesthetic/Economic Composite and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	0.00	0.00
PEOPLE	0.00	0.00
EDUCATION	-0.08	0.01
AGE	0.05	0.00
RESIDENCE	0.02	0.00
UNITS	-0.01	0.00

N = 9,850,000
n = 1,315

Table X-98

Summary Statistics for the Aesthetic/Economic Composite According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.02
Full time	66.7	2.51	3.70	
Part time	4.4	1.55	1.55	
Keeping house	5.9	2.63	3.90	
Retired	20.1	2.64	3.76	
Unemployed	3.0	2.23	2.25	
ETHNICITY				0.00
White	93.0	2.53	3.60	
Nonwhite	7.0	2.25	3.16	
TENURE				0.00
Own	80.6	2.51	3.50	
Rent	15.4	2.24	2.86	
Other	4.0	3.50	6.98	
MOBILE HOME				0.00
Not mobile home	87.8	2.55	3.74	
Mobile home	12.2	2.24	2.62	
PERCEIVED COST				0.00
No cost	4.9	3.21	4.73	
Moderate	88.5	2.47	3.58	
Expensive	5.7	2.34	3.28	
WILLINGNESS TO PAY				0.02
Less	3.5	1.98	2.57	
Same	66.6	2.21	3.16	
More	29.9	3.25	4.49	

N = 9,850,000
n = 1,315

Table X-99

Correlations Between the Aesthetic/Economic Composite and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	0.07	0.01
RELIABILITY	0.11	0.01
ACCESSIBILITY	0.06	0.00
SUPPLY DISTANCE	-0.02	0.00
WELL DEPTH	0.10	0.01
DISTANCE TO CONTAMINATION	0.01	0.00
WELL AGE	0.04	0.00

N = 9,850,000
n = 1,315

Table X-100

Summary Statistics for the Aesthetic/Economic Composite According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.00
Dug	12.7	2.61	4.19	
Bored	6.2	1.92	3.69	
Drilled	67.1	2.57	3.43	
Driven	11.7	2.41	4.15	
Jetted	1.3	1.81	1.74	
Other	1.0	2.90	3.05	
WELL COVER				0.00
Adequate	77.1	2.45	3.43	
Inadequate	22.9	2.70	4.22	
WELL EXIT				0.00
Top of ground	51.5	2.54	3.85	
Pitless adapter	28.9	2.45	2.88	
Buried Seal	18.3	2.51	4.05	
Unclassified	1.4	2.95	2.83	
SPLASHPROOF				0.00
No	21.1	2.78	4.38	
Yes	78.9	2.44	3.39	
GROUTING				0.00
Grouted	42.5	2.56	3.36	
Not grouted	57.5	2.47	3.81	
MAINTENANCE				0.00
Well maintained	78.7	2.40	3.36	
Poorly maintained	21.3	2.92	4.43	
DISPOSAL SYSTEM				0.00
Septic	90.7	2.49	3.59	
Public sewer	3.4	1.97	1.80	
Other	5.9	3.07	4.71	

N = 9,830,000
n = 1,313

Table X-101

Correlations Between the Aesthetic/Economic Composite and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	0.02	0.00
YEARS SERVED	0.02	0.00

N = 9,850,000
n = 1,315

Table X-102

Summary Statistics for the Aesthetic/Economic Composite According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	2.53	3.67	
Public	2.0	2.06	1.15	
Private firm	0.4	1.46	0.50	
Other	0.5	1.25	1.00	
WATER TESTING				0.00
No testing	57.0	2.38	3.48	
Household tests	36.5	2.77	3.98	
System tests, or both household and system tests	6.5	2.19	2.57	

N = 9,850,000
n = 1,315

X-103). Table X-104 shows that the willingness to pay for an improved supply was related to the health-related composite with an eta^2 of 0.07. Households which were willing to pay more for a better supply had water which tended to score slightly worse, on the average, regarding the health-related composite than households which were unwilling to pay more.

Supply attributes

A total of four supply characteristics surpassed the minimum squared association of 0.03. Table X-105 shows total treatment with an R^2 of 0.04, supply distance with an R^2 of 0.03 and well age with an R^2 of 0.03. Table X-106 shows well type to have an eta^2 of 0.07 with the health-related composite. While the average scores on the health-related composite were not greatly different, still, drilled wells had the poorest showing among the classified wells.

System attributes

Among the variables shown in Tables X-107 and X-108, the number of connections had the highest squared association with an R^2 of 0.04. Households on wells which were connected to more households (up to fourteen total) had a slight tendency to have poorer scores on the health-related composite.

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

The variables, education and willingness to pay, combined to account for 7.0 percent of the variance on the health-related composite. Their individual squared associations with the health-related composite were 0.03 and 0.07, respectively. Since, in combination, they have an R^2 of 0.07, there was some overlap in their association with the health-related composite (see Table X-109).

Table X-103

Correlations Between the Health-Related Composite and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	-0.14	0.02
PEOPLE	-0.03	0.00
EDUCATION	0.18	0.03*
AGE	0.06	0.00
RESIDENCE	0.05	0.00
UNITS	0.00	0.00

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-104

Summary Statistics for the Health-Related Composite According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.01
Full time	67.8	30.19	9.78	
Part time	1.9	27.76	0.97	
Keeping house	5.8	28.32	6.45	
Retired	23.8	31.09	10.07	
Unemployed	0.7	32.00	10.00	
ETHNICITY				0.01
White	92.8	30.48	9.74	
Nonwhite	7.2	27.45	6.75	
TENURE				0.00
Own	85.3	30.19	9.22	
Rent	10.2	31.04	11.54	
Other	4.5	29.83	11.31	
MOBILE HOME				0.00
Not mobile home	89.7	30.26	9.51	
Mobile home	11.3	30.30	10.19	
PERCEIVED COST				0.01
No cost	4.4	27.05	3.46	
Moderate	90.9	30.44	9.50	
Expensive	4.7	31.43	14.36	
WILLINGNESS TO PAY				0.07*
Less	2.2	28.49	3.45	
Same	68.1	28.65	8.39	
More	29.7	34.10	11.23	

N = 9,830,000
n = 121

---

*Underscoring denotes that the variable was retained for further analysis.

Table X-105

Correlations Between the Health-Related Composite and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	0.20	0.04*
RELIABILITY	0.03	0.00
ACCESSIBILITY	-0.03	0.00
SUPPLY DISTANCE	0.16	0.03
WELL DEPTH	0.06	0.00
DISTANCE TO CONTAMINATION	0.13	0.02
WELL AGE	0.17	0.03

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-106

Summary Statistics for the Health-Related Composite According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.07*
Dug	14.9	29.20	10.12	
Bored	7.9	27.38	6.12	
Drilled	63.9	31.47	9.93	
Driven	8.8	25.04	4.62	
Jetted	2.2	24.68	2.87	
Other	2.3	38.94	9.51	
WELL COVER				0.02
Adequate	84.0	30.22	9.42	
Inadequate	16.0	30.48	10.40	
WELL EXIT				0.01
Top of ground	46.5	30.79	9.99	
Pitless adapter	25.9	28.91	9.20	
Buried seal	24.6	29.95	8.81	
Unclassified	3.0	36.30	9.52	
SPLASHPROOF				0.01
No	17.2	29.28	10.55	
Yes	82.8	30.47	9.36	
GROUTING				0.01
Grouted	40.1	30.92	10.17	
Not grouted	59.9	29.82	9.15	
MAINTENANCE				0.02
Well maintained	78.6	30.35	9.66	
Poorly maintained	21.4	29.95	9.29	
DISPOSAL SYSTEM				0.00
Septic	92.8	30.46	9.78	
Public sewer	2.2	24.83	5.86	
Other	5.0	28.87	5.82	

N = 9,850,000
n = 121

---

*Underscoring denotes that the variable was retained for further analysis.

Table X-107

Correlations Between the Health-Related Composite and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	0.20	0.04*
YEARS SERVED	0.15	0.02

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-108

Summary Statistics for the Health-Related Composite According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.01
Private groups and individuals	97.0	30.08	9.53	
Public	2.3	37.61	10.53	
Private firm	0.7	32.00	0.00	
Other	0.0	0.00	0.00	
WATER TESTING				0.02
No testing	58.6	30.46	8.96	
Household tests	35.5	29.05	9.62	
System tests, or both household and system tests	5.9	35.60	12.89	

N = 9,850,000
n = 121

Table X-109

Intermediate Regression Estimates for the Health-Related Composite and Household Attributes (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	26.322
	WILLINGNESS TO PAY		
WP_1	Less	b_1	-1.181
WP_2	More	b_2	3.091
X_1	EDUCATION	b_3	0.375

$R^2 = 0.09$

Adjusted $R^2 = 0.07$

Standard error of estimate = 9.16

N = 9,850,000
n = 121

$Y' = a + b_1 WP_1 + b_2 WP_2 + b_3 X_1$

Intermediate supply model

Four independent supply variables, well type, total treatment, supply distance, and well age, were retained for regression analysis. Together, they produced an adjusted R^2 of 0.13 with the health-related composite (see Table X-110).

Intermediate system model

Only one system variable, connections was retained for further analysis. The lack of other independent system variables made an intermediate model unnecessary.

FINAL MODEL SPECIFICATION

The various household, supply, and system variables were combined into a regression model on the health-related composite. Together they produced an adjusted R^2 of 0.39. This was a weak association indicating that the ten variables included were somewhat useful in accounting for the variability of the health-related composite. In general, the health-related composite as an indicator of water quality was not strongly and systematically related across its range with any group of the variables measured (see Table X-111).

General Water Quality Composite

BIVARIATE RELATIONSHIPS

Household attributes

Table X-112 shows that income was weakly related to the general water quality composite with an R^2 of 0.14. The negative correlation indicated the tendency for higher income households to have lower scores, or in other words, to

Table X-110

Intermediate Regression Estimates for the Health-Related Composite and Supply Attributes (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	17.726
	WELL TYPE		
WT_1	Dug	b_1	0.633
WT_2	Bored	b_2	0.119
WT_3	Drilled	b_3	3.013
WT_4	Driven	b_4	-1.606
X_1	TOTAL TREATMENT	b_5	4.101
X_2	SUPPLY DISTANCE	b_6	0.026
X_3	WELL AGE	b_7	0.080

$R^2 = 0.16$

Adjusted $R^2 = 0.13$

Standard error of estimate = 5.70

N = 9,850,000
n = 121

$$Y' = a + b_1WT_1 + b_2WT_2 + b_3WT_3 + b_4WT_4 + b_5X_1 + b_6X_2 + b_7X_3$$

Table X-111

Final Regression Estimates for the Health-Related Composite (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	16.279
	REGION		
R_1	Northeast	b_1	-3.600
R_2	North Central	b_2	3.657
R_3	South	b_3	-3.389
	SMSA/NonSMSA		
S_1	SMSA	b_4	3.654
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	5.452
	WELL TYPE		
WT_1	Dug	b_6	-0.458
WT_2	Bored	b_7	1.261
WT_3	Drilled	b_8	1.917
WT_4	Driven	b_9	-3.581
X_1	CONNECTIONS	b_{10}	4.605
	WILLINGNESS TO PAY		
WP_1	Less	b_{11}	-0.313
WP_2	More	b_{12}	2.761
X_2	TOTAL TREATMENT	b_{13}	2.006
X_3	CONTAMINATION DISTANCE	b_{14}	0.010
X_4	EDUCATION	b_{15}	0.181
X_5	WELL AGE	b_{16}	0.031

Table X-111 continued

$R^2 = 0.44$

Adjusted $R^2 = 0.39$

Standard error of estimate = 7.12

N = 9,850,000
n = 121

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6WT_1 + b_7WT_2 + b_8WT_3 + b_9WT_4 + b_{10}X_1 + b_{11}WP_1 + b_{12}WP_2 + b_{13}X_2 + b_{14}X_3 + b_{15}X_4 + b_{16}X_5$$

Table X-112

Correlations Between the General Water Quality Composite and Continuous Household Variables (For Households Served by Wells)

Variable	r	R^2
INCOME	-0.25	0.06*
PEOPLE	0.01	0.00
EDUCATION	-0.04	0.00
AGE	0.11	0.00
RESIDENCE	0.12	0.01
UNITS	0.00	0.00

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

have better general water quality. Households which expressed a willingness to pay more for a better water supply tended to have poorer quality water. The overall association was an eta^2 of 0.08 (see Table X-113).

Supply attributes

Three continuous and three categorical supply variables had squared measures of association at or above the 0.03 criterion. Table X-114 shows total treatment with an R^2 of 0.07, well age with an R^2 of 0.04, and reliability with an R^2 of 0.03. The correlations indicated that households with more treatment devices connected to the well serving them tended to have poorer quality water. Though the NSA had no way of assessing it, this probably reflected the condition that those supplies with more treatment devices started out with poor quality source water. Less reliable and older wells showed a slight tendency to have poorer quality water.

Table X-115 shows that grouting, well type, and well exit were each at or above the 0.03 criterion of statistical significance. They had squared associations with the general water quality composite of 0.05, 0.04, and 0.03, respectively.

System attributes

None of the system characteristics had associations with the general water quality composite greater than 0.02 (see Tables X-116 and X-117).

INTERMEDIATE MODEL SPECIFICATION

Intermediate household model

The combination of willingness to pay and total household income showed an adjusted R^2 of 0.12 with the general water quality composite (see Table X-118).

Table X-113

Summary Statistics for the General Water Quality Composite According to Categorical Household Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WORKING				0.00
Full time	67.8	9.67	8.15	
Part time	1.9	11.09	2.43	
Keeping house	5.8	9.06	8.24	
Retired	23.8	10.04	9.25	
Unemployed	0.7	10.00	0.00	
ETHNICITY				0.01
White	92.8	9.99	8.54	
Nonwhite	7.2	6.65	4.08	
TENURE				0.00
Own	85.3	9.95	8.55	
Rent	10.2	8.02	6.32	
Other	4.5	9.90	7.91	
MOBILE HOME				0.00
Not mobile home	88.7	9.70	8.06	
Mobile home	11.3	10.15	10.27	
PERCEIVED COST				0.00
No cost	4.4	9.75	7.46	
Moderate	90.9	9.80	8.16	
Expensive	4.7	9.55	12.65	
WILLINGNESS TO PAY				0.08*
Less	2.2	7.96	2.90	
Same	68.1	8.28	6.77	
More	29.7	13.26	10.54	

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-114

Correlations Between the General Water Quality Composite and Continuous Supply Variables (For Households Served by Wells)

Variable	r	R^2
TOTAL TREATMENT	0.26	0.07*
RELIABILITY	0.17	0.03
ACCESSIBILITY	-0.03	0.00
SUPPLY DISTANCE	0.11	0.01
WELL DEPTH	-0.11	0.01
DISTANCE TO CONTAMINATION	0.15	0.02
WELL AGE	0.21	0.04

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-115

Summary Statistics for the General Water Quality Composite According to Categorical Supply Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.04*
Dug	14.9	11.79	11.40	
Bored	7.9	5.46	3.17	
Drilled	63.9	10.05	8.24	
Driven	8.8	7.83	5.00	
Jetted	2.2	6.03	4.33	
Other	2.3	13.82	3.28	
WELL COVER				0.00
Adequate	84.0	9.11	7.97	
Inadequate	16.0	13.13	9.35	
WELL EXIT				0.03
Top of ground	46.5	11.11	9.70	
Pitless adapter	25.9	7.51	4.65	
Buried seal	24.6	8.95	8.29	
Unclassified	3.0	14.58	3.17	
SPLASHPROOF				0.01
No	17.2	11.13	9.04	
Yes	82.8	9.46	8.16	
GROUTING				0.05
Grouted	40.1	10.25	8.78	
Not grouted	59.9	9.42	8.02	
MAINTENANCE				0.02
Well maintained	78.6	9.18	7.87	
Poorly maintained	21.4	11.85	9.60	
DISPOSAL SYSTEM				0.00
Septic	92.8	9.84	8.53	
Public sewer	2.2	9.48	2.65	
Other	5.0	8.15	5.79	

N = 9,850,000
n = 121

*Underscoring denotes that the variable was retained for further analysis.

Table X-116

Correlations Between the General Water Quality Composite and Continuous System Variables (For Households Served by Wells)

Variable	r	R^2
CONNECTIONS	0.12	0.01
YEARS SERVED	0.13	0.02

N = 9,850,000
n = 121

Table X-117

Summary Statistics for the General Water Quality Composite According to Categorical System Variables (For Households Served by Wells)

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.00
Private groups and individuals	97.1	9.68	8.43	
Public	2.0	12.83	4.46	
Private firm	0.7	10.00	0.00	
Other	0.0	0.0	0.0	
WATER TESTING				0.01
No testing	58.6	9.09	6.75	
Household tests	35.5	10.29	10.08	
System tests, or both household and system tests	5.9	12.98	9.90	

N = 9,850,000
n = 121

Table X-118

Intermediate Regression Estimates for the General Water Quality Composite and Household Attributes (For Households Served by Wells)

Variable	Designation	Parameter	Estimate
	Intercept	a	11.885
	WILLINGNESS TO PAY		
WP_1	Less	b_1	-2.585
WP_2	More	b_2	3.674
X_1	INCOME	b_3	0.000

$R^2 = 0.13$

Adjusted $R^2 = 0.12$

Standard error of estimate = 7.31

N = 9,850,000
n = 121

$$Y' = a + b_1 WP_1 + b_2 WP_2 + b_3 X_1$$

Intermediate supply model

The supply variables retained for model development included total treatment, well age, grouting, well type, reliability, and well exit. Combined, they showed an R^2 of 0.25 with the general water quality composite (see Table X-119). In general, there was a weak relationship between the general character of the household water and the particular well configuration employed. It was expected that if the quality of the source water could have been controlled in the analysis, this association would have been stronger.

Intermediate system model

None of the system characteristics were related to the general water quality composite sufficiently to warrant modeling.

FINAL MODEL SPECIFICATION

Table X-120 demonstrates that the household, supply, and sample design attributes combine to resolve 40.0 percent of the general water quality composite. This demonstrated association is sufficient to deduce that geographic location, some household and some supply attributes are meaningfully related to the general quality of water. However, the association is moderate at best. Perhaps, as mentioned earlier, knowledge of the source water quality would clarify these relationships.

DISCUSSION

None of the associations of household, supply, or system characteristics with the various water quality indicators was very strong. The most persistent variable, related to several of the water quality indicators, was the well type—the type of construction. The strongest overall associations were found with the health-related and the general water quality composite. But, the number of cases

Table X-119

Intermediate Regression Estimates for the General Water Quality Composite and Supply Attributes (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	-4.020
X_1	TOTAL TREATMENT	b_1	4.399
X_2	WELL AGE	b_2	0.110
G_1	GROUTING	b_3	1.572
	WELL TYPE		
WT_1	Dug	b_4	2.845
WT_2	Bored	b_5	-3.139
WT_3	Drilled	b_6	1.101
WT_4	Driven	b_7	0.608
X_3	RELIABILITY	b_8	0.893
	WELL EXIT		
WE_1	Top of ground	b_9	2.433
WE_2	Pit	b_{10}	-1.325

$R^2 = 0.29$

Adjusted $R^2 = 0.25$

Standard error of estimate = 7.10

N = 9,850,000
n = 121

$$Y' = a + b_1X_1 + b_2X_2 + b_3G_1 + b_4WT_1 + b_5WT_2 + b_6WT_3 + b_7WT_4 + b_8X_3 + b_9WE_1 + b_{10}WE_2$$

Table X-120

Final Regression Estimates for the General Water Quality Composite (For Households Served by Wells)

Variable Designation		Parameter	Estimate
	Intercept	a	1.997
	REGION		
R_1	Northeast	b_1	-0.665
R_2	North Central	b_2	1.483
R_3	South	b_3	-2.788
	SMSA/NonSMSA		
S_1	SMSA	b_4	2.069
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-1.219
	WILLINGNESS TO PAY		
WP_1	Less	b_6	-1.226
WP_2	More	b_7	2.176
X_1	TOTAL TREATMENT	b_8	4.002
X_2	INCOME	b_9	0.000
X_3	WELL AGE	b_{10}	0.073
G_1	GROUTING	b_{11}	1.101
	WELL TYPE		
WT_1	Dug	b_{12}	1.707
WT_2	Bored	b_{13}	-1.251
WT_3	Drilled	b_{14}	0.869
WT_4	Driven	b_{15}	-1.714
	WELL EXIT		
WE_1	Top of ground	b_{16}	2.364
WE_2	Pit	b_{17}	-2.065
X_4	RELIABILITY	b_{18}	0.727

Table X-120 continued

$R^2 = 0.45$

Adjusted $R^2 = 0.40$

Standard error of estimate = 6.21

N = 9,850,000
n = 121

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6WP_1 + b_7WP_2 + b_8X_1 + b_9X_2 + b_{10}X_3 + b_{11}G_1 + b_{12}WT_1 + b_{13}WT_2 + b_{14}WT_3 + b_{15}WT_4 + b_{16}WE_1 + b_{17}WE_2 + b_{18}X_4$$

available for each of these prohibits strong statements regarding the national condition. At least in part, the lack of associations demonstrated were probably due to relationships which were more complex than could be usefully reduced to a one number measure of linear association. In the succeeding approach, total coliform is examined with contingency tables to allow some of the complexities of association to be manifest more fruitfully than could be done with regression.

TOTAL COLIFORM AMONG ALL RURAL HOUSEHOLDS: A CONTINGENCY TABLE APPROACH

A majority of rural US households were served by supplies which, according to NSA criteria for any single water quality parameter, were safe for human consumption. Therefore, the foregoing NSA investigation of overall national, linear, systematic associations with water quality was strongly affected by the fact that most supplies were acceptable. The effect was to force most squared associations toward zero.

Most rural households were served by groundwater. The well technology in use frequently involved pumping and piping the water without further treatment. Therefore, households in areas where the groundwater tended to be safe usually had safe water, regardless of other household, supply, or system characteristics. On the other hand, in those areas where the groundwater had a problem, the problem appeared in the water of households of many types, served by supplies of wide variety. But, when a problem was present, some household, supply, and system attributes did appear related to the susceptibility of the household water to be affected. In the foregoing analysis, it was difficult to highlight these latter relationships. Therefore, a more extended analysis was undertaken here to better specify the conditions which were related to a problem. In this analysis, contingency tables are used to see if there are subsets of the data which show important relations with total coliform contamination that were not apparent in the regression analyses.

Specifically, an extended analysis was made of coliform incidence to identify attributes of households which were related to water quality problems, as well as those which were notably unrelated.

For this analysis, coliform counts were divided into two groups. The first group had households with one or fewer total coliform organisms per 100 milliliters of water (the NSA coliform reference value). The second group had households with more than one coliform per 100 milliliters of water. This second group was further divided into a 2 through 99 and a 100-plus count group. Similarly, each of the continuous household, supply, and system characteristics were categorized. Table X-121 lists the variables and the categorizations which were chosen. The NSA reference values described in Chapters V and VI were used to delineate some categories. Arbitrary breakpoints were chosen at other points of the distributions. Different categorizations could have suggested alternative interpretations.

Not all of the household, supply, and system characteristics are presented for each water quality indicator; rather, only those which demonstrated patterns that suggested fairly substantial differences in stress among households are presented.

Total Coliform

Throughout rural America, 71.1 percent of all households had one or fewer total coliforms per 100 milliliters of water. Consequently, the incidence of total coliform contamination had a fairly low linear association with the variables considered, with the greatest association between the type of supply and the coliform

Table X-121

Categories Chosen for the Water Quality Indicators and the Continuous Household, Supply, and System Attributes

Variable	Categories
INCOME	Less than $10,000 $10,000 - 19,999 $20,000 - 29,999 $30,000 or more
PEOPLE	1 person 2 people 3 people 4 people 5 or more people
EDUCATION	Less than high school Completed high school Schooling beyond high school
AGE	Less than 25 years 25 - 49 years 50 - 64 years 65 or more
RESIDENCE	1 year or less 2 - 4 years 5 - 9 years 10 - 19 years 20 or more
UNITS	Single family dwelling Multiple family dwelling
TOTAL TREATMENT	Count of the number of treatment devices from 0 to 8
RELIABILITY	Good (<2.0) Medium (2.0 - 3.9) Poor (≥4.0)
ACCESSIBILITY	Good (<8.0) Medium (8.0 - 8.9) Poor (≥9.0)

Table X-121 continued

Variable	Categories
CONNECTIONS	1 connection 2 - 14 connections 15 or more connections
YEARS SERVED	0 - 4 years 5 - 19 years 10 - 19 years 20 or more years

counts. (Wells tended to have higher counts than community supplies.) The eta^2 between type of supply and coliform counts was 0.16 while, with all of the relevant variables combined, the R^2 of the final model was 0.19, less than one-fifth of the total variance.

Using the same independent variables, important relationships were found with coliform counts which were not visible in the regression analyses. The regression analyses summarize associations across the entire range of the data, whereas contingency tables, on the other hand, can depict aberrations in the association within the range. Therefore, coliform counts were reexamined using a contingency table approach.

Not all of the variables listed in Table X-121 or originally shown in Table X-3 showed any different association or interpretation from that presented earlier in this chapter. Those characteristics included: the number of people in the household, the age of the household head, the length of residence, the number of dwelling units in the building, the employment status of the head, the ethnicity of the head, household tenure, the number of water treatment devices, the reliability of the supply, and the water testing history. The remaining variables had associations which merit further discussion.

Income

The association between the combined income of all household members and the count of coliforms was very low, slightly more than an R^2 of 0.01. Table X-122 shows four categories of income against breakdowns of coliform counts. The low R^2 is understandable since a majority of households in each of the income categories had acceptable water (one or fewer total coliforms per 100 milliliters of water). But, the pattern of households with total coliform problems was importantly different as income increases. Among low income households, which reported less than $10,000 total income, 65.2 percent had acceptable total coliform

Table X-122

Percent of Households with Various Degrees of Total Coliform Contamination for Each Income Level

Coliform Counts	Income Levels				Number and Percent of Households
	$10,000	$10,000 - 19,999	$20,000 - 29,999	$30,000 +	
Households with one or fewer coliforms per 100 milliliters of water	65.2	70.1	73.3	83.3	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	34.7	29.9	26.7	16.8	6,350,000 28.9
2 - 99 coliforms	18.4	19.5	19.2	12.3	3,960,000 18.0
100 + coliforms	16.3	10.4	7.5	4.5	2,390,000 10.9
Number of Households	6,750,000	8,090,000	3,970,000	3,170,000	21,970,000
Percent of Total	30.7	36.8	18.1	14.4	

readings. The higher income group, reporting $30,000 or more, had 83.3 percent with one or fewer coliforms, a relative difference of 18.1 percent. While most households in each income group were likely to have a relatively coliform-free supply, Table X-122 showed that wealthier households were substantially more likely to have acceptable supplies than low income households. Low income households were over three times more likely to have high coliform counts (100 or more total coliforms per 100 milliliters of water) than the upper income households (16.3 percent versus 4.5 percent). While most households of every income level had acceptable total coliform levels, the risk of a coliform problem tended to be less among higher income households.

Education

A similar, though not as striking, pattern was manifest between coliform and education (see Table X-123). Households with heads that had less than a high school education were more likely to experience coliform problems than households with heads who were more highly educated. Over 35 percent of the low education groups had total coliform counts above the NSA reference value of one coliform per 100 milliliters of water; among households of the high education group, that percentage dropped to 20.8. Taken together, income and education comprise important components of socioeconomic status. More than half of the rural households in the groups considered had acceptable coliform levels, but high coliform counts were notably more common among households with lower socioeconomic status.

Perceived cost

Rural residents were asked to evaluate the cost of their water, whether they thought it was relatively inexpensive or very expensive. A few households, 4.4 percent of the total, reported that their water was free. A slightly smaller

Table X-123

Percent of Households with Various Degrees of Total Coliform Contamination for Three Levels of Education

Coliform Counts	Education Level			Number and Percent of Households
	Less Than High School	High School	More Than High School	
Households with one or fewer coliforms per 100 milliliters of water	64.7	69.3	79.2	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	35.3	30.7	20.8	6,350,000 28.9
2 - 99 coliforms	20.4	19.7	14.0	3,960,000 18.0
100 + coliforms	14.9	11.0	6.7	2,390,000 10.9
Number of Households	7,740,000	6,810,000	7,420,000	21,970,000
Percent of Total	35.2	31.0	33.8	

proportion, 3.5 percent, thought their supply was very expensive. While the proportion of households in these extreme categories was small, they marked the endpoints of a definite pattern. Households which reported increasingly expensive water supplies were less likely to have coliform problems (see Table X-124). Over half (57.8 percent) of the households which reported no cost for their water had coliform counts above the NSA reference value, while only 7.0 percent of the households which reported their supply was very expensive were above the reference value.

Willingness to pay

Household residents were asked if they would be willing to pay more for an ideal supply than they were paying for their current supply. About one-third said they would pay more for such a supply, 3.6 percent said they would pay less than their present cost for such a supply, and 63.3 percent said they would pay the same as their present arrangement. Table X-125 shows that households which reported they would pay more to have an improved supply were more likely to have a coliform problem (32.2 percent, compared to 24.6 percent for households willing to pay less).

Both willingness to pay and perceived cost were householder evaluations of the water. Apparently many residents of households with extremely good water and very poor water evaluated their supply consistent with the laboratory indicators of quality. Households willing to pay more for a better supply were giving an evaluation that the current supply was deficient in some way. The laboratory measures of total coliform tended to confirm the perceived deficiency.

Accessibility

Table X-126 relates the degree of supply accessibility to total coliform counts. Less accessible supplies were more likely to have high coliform counts. In

Table X-124

Percent of Households with Various Degrees of Total Coliform Contamination by Perceived Cost

	Perceived Costs					Number and Percent of Households
	No Cost	Inexpensive	Reasonable	Expensive	Very Expensive	
Households with one or fewer coliforms per 100 milliliters of water	42.2	63.5	73.7	81.5	93.0	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	57.8	36.3	26.6	18.4	7.0	6,350,000 28.9
2 - 99 coliforms	28.0	20.5	17.9	12.8	7.0	3,960,000 18.0
100 + coliforms	29.8	15.8	8.7	5.6	0.0	2,390,000 10.9
Number of Households	961,000	5,840,000	11,900,000	2,480,000	761,000	21,970,000
Percent of Total	4.4	26.6	54.2	11.3	3.5	

Table X-125

Percent of Households with Various Degrees of Total Coliform Contamination for Each Level of Willingness to Pay

Coliform Counts	Willingness to Pay			Number and Percent of Households
	Less	Same	More	
Households with one or fewer coliforms per 100 milliliters of water	75.4	72.7	67.8	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	24.6	27.5	32.2	6,350,000 28.9
2 - 99 coliforms	14.1	18.1	18.5	3,960,000 18.0
100 + coliforms	10.5	9.4	13.7	2,390,000 10.9
Number of Households	780,000	13,900,000	7,270,000	21,970,000
Percent of Total	3.6	63.3	33.1	

Table X-126

Percent of Households with Various Degrees of Total Coliform Contamination for Three Categories of Accessibility

Coliform Counts	Accessibility Level			Number and Percent of Households
	Good	Medium	Poor	
Households with one or fewer coliforms per 100 milliliters of water	76.9	64.6	40.4	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	23.2	35.4	59.9	6,350,000 28.9
2 - 99 coliforms	16.2	20.1	28.8	3,960,000 18.0
100 + coliforms	7.1	15.3	31.1	2,390,000 10.9
Number of Households	14,300,000	6,240,000	1,380,000	21,970,000
Percent of Total	65.1	28.4	6.3	

fact, over half (59.9 percent) of the approximately 1.4 million households with poorly accessible supplies had counts above the NSA reference value. Moreover, 31.1 percent of these households registered counts of 100 or more coliforms per 100 milliliters of water. Of those households with highly accessible supplies, only 23.2 percent were above the NSA reference value and about 7 percent registered counts of 100 or more coliforms per 100 milliliters of water.

Type of supply

Table X-127 shows three categories of supply: community, well, and other. Elsewhere in the report, the distinction was made among community, intermediate, and individual systems. In this instance, community still refers to households served by systems having fifteen or more connections. But, intermediates and individuals are divided differently according to whether the supply is a well (with fewer than fifteen connections) or some other arrangement. "Other" supplies have fewer than 15 connections and include surface water, springs, cisterns, hauled, and purchased bottled water. The great majority of households served by community and well supplies had acceptable total coliform counts, with community supplies substantially more likely to have acceptable coliform counts (84.4 percent, compared to 61.7 percent of wells with coliform counts below the NSA reference value.) On the other hand, more than three-quarters of the 1.1 million households served by "other" supplies had coliform counts above the NSA reference value, compared to 15.5 percent of community supplies. Though the eta^2 measure of association between supply type and coliform counts was only 0.16, because so many households had acceptable counts, Table X-127 clearly shows that the category of supply serving a household is a valuable indicator of the relative likelihood of coliform contamination.

Table X-127

Percent of Households with Various Degrees of Total Coliform Contamination for Each Type of Supply

Coliform Counts	Supply Type			Number and Percent of Households
	Community Water Supply	Well	Other	
Households with one or fewer coliforms per 100 milliliters of water	84.4	61.7	22.5	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	15.5	38.2	77.7	6,350,000 28.9
2 - 99 coliforms	12.4	22.9	28.8	3,960,000 18.0
100 + coliforms	3.0	15.2	48.9	2,390,000 10.9
Number of Households	11,000,000	9,850,000	1,140,000	21,970,000
Percent of Total	50.1	44.8	5.2	

Number of system connections

Most single-connection systems were wells. All of the systems categorized in the NSA as community systems had 15 or more connections. Thus, Table X-128 duplicates much of the information shown in Table X-127, and shows that systems with more connections are less likely to have high coliform counts. Only 3.0 percent of the fifteen or more connection systems had coliform counts at or above 100 organisms per 100 milliliters of water. On the other hand, 20.0 percent of the one-connection systems fell in the high count category.

Years served by present supply

Residents at each household were asked how many years they had been served by the current major household water supply. The responses reflected two factors. First, if the residents obtained a new supply during their tenure, then the response really was the length of time the "new" supply had been functioning for that household. Second, if the current supply was already serving the household when the residents moved in, then the response was a reflection of how long the current residents had inhabited the dwelling.

Table X-129 displays coliform counts by four categories for the number of years served. About 66 percent of all rural US households fell in the first two categories of zero to four years and five to nine years. Among the shorter length of service categories, the more recent systems, or the households that had been recently occupied, tended to demonstrate lower coliform counts. Among the latter two categories of longer length residents, newer systems also tended to have lower counts.

Table X-129 could not be clearly defined to reflect the length of supply service from the length of residence. It was, therefore, difficult to draw definitive conclusions. The break in pattern between the five to nine and the ten to nineteen year groups and the almost duplicate pattern between the two low categories and

Table X-128

Percent of Households with Various Degrees of Total Coliform Contamination by the Number of System Connections

Coliform Counts	Number of Connections 1	2 - 4	15 +	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	57.9	56.8	84.4	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	42.1	43.5	15.5	6,350,000 28.9
2 - 99 coliforms	22.1	29.7	12.4	3,960,000 18.0
100 + coliforms	20.0	13.8	3.0	2,390,000 10.9
Number of Households	8,770,000	2,220,000	11,000,000	21,970,000
Percent of Total	39.9	10.1	50.1	

Table X-129

Percent of Households with Various Degrees of Total Coliform Contamination by the Number of Years Served

Coliform Counts	Number of Years Served				Number and Percent of Households
	0 - 4	5 - 9	10 - 19	20 +	
Households with one or fewer coliforms per 100 milliliters of water	73.5	65.7	73.6	67.3	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	26.2	34.0	26.4	32.7	6,350,000 28.9
2 - 99 coliforms	17.6	22.1	16.9	15.5	3,960,000 18.0
100 + coliforms	8.6	12.0	9.6	17.2	2,390,000 10.9
Number of Households	10,300,000	4,290,000	3,750,000	3,670,000	21,970,000
Percent of Total	46.9	19.5	17.1	16.7	

the two high categories of service length were probably a function of the length of residence as explained above, but that was not certain.

System ownership

Table X-130 shows differences in patterns of total coliform contamination according to the type of system ownership. Households served by systems owned publicly or by a private firm had fewer coliform problems than those owned by a private individual or "other" types of ownership. Between 15 and 20 percent of the former group's households were above the NSA reference value. By contrast, over 40 percent of the latter group's households were above the reference value. Publicly owned systems, at 2.7 percent, were the least likely to have high counts (100 or more coliforms per 100 milliliters) while 17.9 percent of the systems owned by a private individual were likely to have households in that high category.

Discussion

Several of the characteristics of the system, when joined, yielded an especially strong relationship to the incidence of over-reference value coliform. Households on systems which were privately owned, which had fewer than fifteen connections, which were judged to be poorly accessible, or any combination of the three, made up just over half of all households (54.1 percent). But that same group of households included 75.6 percent of all over-coliform-reference value households in the NSA.

Table X-131 shows that of the 45.9 percent of NSA households served by systems with none of those selected system characteristics, 15.4 percent had over-reference value coliform counts. At the other extreme, among the 3.4 percent of NSA households served by systems with all three characteristics, 63.5 percent had coliform counts over the NSA reference value.

Table X-130

Percent of Households with Various Degrees of Total Coliform Contamination for Types of System Ownership

Coliform Counts	System Ownership				Number and Percent of Households
	Private	Public	Firm	Other	
Households with one or fewer coliforms per 100 milliliters of water	59.2	84.8	80.3	53.0	15,620,000 71.1
Households with more than one coliform per 100 milliliters of water	40.6	15.2	19.3	47.0	6,350,000 28.9
2 - 99 coliforms	22.7	12.5	13.1	38.6	3,960,000 18.0
100 + coliforms	17.9	2.7	6.2	8.4	2,390,000 10.9
Number of Households	11,500,000	8,950,000	1,470,000	76,874	21,970,000
Percent of Total	52.3	40.7	6.7	0.3	

Table X-131

Relationship Between the Count of Selected System Characteristics* and the Probability of a Total Coliform Problem Among Rural Households

Count of Specified Characteristics	Percent of Rural Households	Probability of a Coliform Problem (more than 1 coliform per 100 milliliters of water)
With 0 characteristics	45.9	15.4
With 1 characteristic	6.3	27.8
With 2 characteristics	44.4	40.6
With 3 characteristics	3.4	63.5

N = 21,970,000
n = 2,654

*Selected system characteristics were: system had fewer than fifteen connections; system was privately owned; system was judged to be poorly accessible.

TOTAL COLIFORM AMONG RURAL HOUSEHOLDS SERVED BY WELLS: A CONTINGENCY TABLE APPROACH

There were 9.8 million rural households served by wells. Fully 38.2 percent of these had coliform counts in excess of the NSA coliform reference value of one coliform organism per 100 milliliters of water. The following analysis parallels the previous presentation on total coliform contamination among all rural households. However, for this special analysis of coliform counts in well supplies, several variables on the construction, siting, and maintenance of wells were added. Table X-132 describes the additional variables.

Several household, supply, and system characteristics did not show meaningful associations with the coliform counts for various reasons. They included the number of people living in the household, the age of the household head, the length of residence, the working status of the household head, the ethnicity of the household head, and the number of dwelling units in the building. In some cases, one category of the variable contained a preponderance of cases. For instance, almost 99 percent of rural households served by noncommunity system wells were single family dwelling units. In other cases, the differences in total coliform counts among categories were not sufficient to warrant interpretation. Similarly, none of these variables showed a meaningful relationship with total coliform contamination among all rural households regardless of supply.

On the other hand, some of the variables which did not merit discussion in the analysis of coliform contamination for the nation as a whole, did show differences among households served by wells in the contingency table analysis. These included: tenure, devices, reliability, and water test history.

Table X-132
Description of Well Attributes

Name	Description	Discussion
WELL TYPE	Type of Well Construction Dug Bored Drilled Driven Jetted Other	VII
GROUTING	Was the well grouted No Yes	VII
WELL EXIT	How the water line leaves the well Pitless adapter Buried seal Neither (well terminates above ground) Unclassified	VII
SPLASH PROOF	Did the well cover appear to be splashproof? No Yes Unclassified	VII
COVER	Did the well cover appear to be adequate to protect against intrusion (animals, insects, etc.) Inadequate Adequate Unclassified	VII
MAINTENANCE	Did the well appear to be adequately maintained? Inadequate Adequate	VII
DEPTH	The depth of the well 3 meters or less 3.1 - 9.9 meters 10 - 14.9 meters 15 - 19.9 meters 20 - 29.9 meters 30 - 59.9 meters 60 + meters	

Table X-132 continued

Name	Description	Discussion
WELL AGE	The age of the well 5 years or less 5.1 - 10 years 10.1 - 20 years 20 + years	VII
CONTAMINATION DISTANCE	Distance to the nearest visible potential contamination source 0 - 5 meters 6 - 10 meters 11 - 15 meters 15 + meters	VII
WASTEWATER DISPOSAL	General category of wastewater disposal Unconfined (dump on ground, direct discharge to water body, privy, etc.) Septic Public sewer	VII

Income

Among households connected to wells with fewer than fifteen connections, there was a consistent pattern that households in higher income categories were less likely to have coliform bacteria present in their water. Furthermore, the proportion of households in Table X-133 with coliform counts greater than 100 was about three times greater among households which reported income under $10,000 than among households above $30,000. The basic pattern was almost identical to that for all rural households, except about 9 percent more households served by wells had detectable total coliform bacteria than rural households in general. Most households of every income category had fewer than two coliforms per 100 milliliters of water ranging from 55.7 percent among households earning less than $10,000 per year to 74.2 percent among the $30,000 and up category.

Education

Table X-134 shows that households where the head had a higher education tended to have coliform problems less often than households where the head had less formal education. As with income, the pattern was similar to that seen earlier for all rural households in general. Considered jointly with income, it appeared that higher socioeconomic status was associated with lower likelihood of coliform contamination.

Mobile home

Table X-135 indicates that wells serving mobile homes, as opposed to other dwelling types (single family houses being the most common), more frequently had coliform contamination. Among mobile homes, 46.0 percent had water which tested as having more than one coliform per 100 milliliters, compared to 37.1 percent of the nonmobile home dwellings.

Table X-133

Percent of Households Served by Wells with Various Degrees of Coliform Contamination for Each Income Level

Coliform Counts	Income Levels				Number and Percent of Households
	$10,000	$10,000 - 19,999	$20,000 - 29,999	$30,000 +	
Households with one or fewer coliforms per 100 milliliters of water	55.7	61.7	64.7	74.2	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	44.3	38.3	35.1	25.7	3,765,000 38.2
2 - 99 coliforms	22.1	24.6	24.4	18.2	2,264,000 23.0
100 + coliforms	22.2	13.7	10.9	7.5	1,501,000 15.2
Number of Households	3,244,000	3,646,000	1,798,000	1,161,000	9,850,000
Percent of Total	32.9	37.0	18.3	11.8	100.0

Table X-134

Percent of Households Served by Wells with Various Degrees of Coliform Contamination for Three Levels of Education

Coliform Counts	Education Level: Less Than High School	Education Level: High School	Education Level: More Than High School	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	57.0	58.8	73.0	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	43.0	41.2	27.0	3,765,000 38.2
2 - 99 coliforms	24.9	25.0	17.4	2,264,000 23.0
100 + coliforms	18.1	16.1	9.6	1,501,000 15.2
Number of Households	3,948,000	3,339,000	2,563,000	9,850,000
Percent of Total	40.1	33.9	26.0	100.0

Table X-135

Percent of Mobile Homes and Nonmobile Homes Served by Wells with Various Degrees of Coliform Contamination

Coliform Counts	Not Mobile Home	Mobile Home	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	62.9	54.0	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	37.1	46.0	3,765,000 38.2
2 - 99 coliforms	21.9	30.7	2,264,000 23.0
100 + coliforms	15.2	15.3	1,501,000 15.2
Number of Households	8,651,000	11,988,000	9,850,000
Percent of Total	87.8	12.2	100.0

Tenure

Rural residents who own their home had more than one coliform per 100 milliliters of their water in 37.0 percent of the cases (Table X-136). Renters, on the other hand, were more likely to have a problem. Over 47 percent of renters had comparable coliform counts. A small proportion, 4.0 percent, of rural households are nonstandard tenure situations such as children who live in a separate house owned by a parent but do not pay rent. Unaccountably, this small group was the least likely of the three to display a coliform problem. The difference was most apparent among the 2-99 coliform count category. There was little difference among the three tenure categories regarding the proportions of households with 100 or more coliforms per 100 milliliters of water.

Perceived cost

Table X-137 suggests a pattern that supplies perceived to be more expensive tended to have lower rates of total coliform incidence. But, the low counts of households in the extreme categories made that pattern tenuous. The pattern was similar to that which was seen for households in general, except that households served by wells tended to cluster more into the 'inexpensive' and 'reasonable' categories.

Willingness to pay

Most residents in rural households served by wells said they would not be willing to pay any more for a hypothetically ideal supply than they paid for their present supply. But, it was the case that those who said they would pay less were found to experience coliform contamination less often than other households. On the other hand, households which reported a willingness to pay more in order to secure the better supply, were above the NSA coliform reference value more often.

Table X-136

Percent of Households Served by Wells with Various Degrees of Coliform Contamination for Households of Various Tenure Types

Coliform Counts	Type of Tenure			Number and Percent of Households
	Own	Rent	Other	
Households with one or fewer coliforms per 100 milliliters of water	63.0	52.8	72.5	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	37.0	47.3	27.6	3,765,000 38.2
2 - 99 coliforms	22.0	30.7	13.1	2,264,000 23.0
100 + coliforms	15.0	16.6	14.5	1,501,000 15.2
Number of Households	7,937,000	1,519,000	394,000	9,850,000
Percent of Total	80.6	15.4	4.0	100.0

Table X-137

Percent of Households Served by Wells with Various Degrees of Coliform Contamination Reporting Various Levels of Perceived Cost for Their Water

Coliform Counts	Perceived Costs					Number and Percent of Households
	No Cost	Inexpensive	Reasonable	Expensive	Very Expensive	
Households with one or fewer coliforms per 100 milliliters of water	52.6	61.8	61.3	67.9	95.9	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	47.4	38.2	38.7	32.1	4.1	3,765,000 38.2
2 - 99 coliforms	32.1	21.6	23.8	20.6	4.1	2,264,000 23.0
100 + coliforms	15.3	16.5	14.9	11.5	0.0	1,501,000 15.2
Number of Households	486,000	3,962,000	4,754,000	560,000	88,000	9,850,000
Percent of Total	4.9	40.2	48.3	5.7	0.9	100.0

This pattern, as shown in Table X-138, was similar to the pattern for households served by all types of supply (see Table X-125).

Accessibility

Very few wells (3.6 percent) were judged to be poorly accessible supplies. But, of those so judged, 58.8 percent failed the NSA reference value of one coliform organism per 100 milliliters of water. That rate, as shown in Table X-139, was almost identical to the failure rate among poorly accessible supplies throughout the nation. Different from the national picture, which showed the most accessible supplies to have the lowest rate of coliform problems, wells judged to have medium and good accessibility showed negligible differences of total coliform problems.

Reliability

Table X-140 indicates that the 18.2 percent of households served by wells were judged to be moderately or poorly reliable. In both instances, almost half of the households exceeded the NSA coliform reference value. On the other hand, households served by wells judged to have good reliability failed about 36 percent of the time. Reliability was determined by the supply's history of breakdowns. Thus the frequency of breakdowns and was positively associated with avenues for microbial contamination.

Total treatment

Table X-141 displays the rates of coliform problems among households served by wells which had varying numbers of water treatment devices connected to them. More than half of those wells with no treatment device of any type had counts greater than one coliform per 100 milliliters of water. As the number of treatment devices (filter, chlorinators, etc.) increased, the incidence of coliform

Table X-138

Percent of Households Served by Wells with Various Degrees of Coliform Contamination with Varying Degrees of Willingness to Pay for an Improved Supply

Coliform Counts	Willing to Pay			Number and Percent of Households
	Less	Same	More	
Households with one or fewer coliforms per 100 milliliters of water	69.5	62.9	58.3	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	30.5	37.1	41.7	3,765,000 38.2
2 - 99 coliforms	12.9	24.2	21.5	2,264,000 23.0
100 + coliforms	17.7	12.9	20.2	1,501,000 15.2
Number of Households	349,000	6,561,000	2,940,000	9,850,000
Percent of Total	3.5	66.6	29.9	100.0

Table X-139

Percent of Households With Varying Accessibility Served by Wells with Various Degrees of Coliform Contamination

Coliform Counts	Accessibility			Number and Percent of Households
	Good	Medium	Poor	
Households with one or fewer coliforms per 100 milliliters of water	62.5	62.8	41.2	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	37.5	37.2	58.8	3,765,000 38.2
2 - 99 coliforms	23.4	20.4	28.8	2,264,000 23.0
100 + coliforms	14.1	16.8	30.0	1,501,000 15.2
Number of Households	7,496,000	2,000,000	354,000	9,850,000
Percent of Total	76.1	20.3	3.6	100.0

Table X-140

Percent of Households With Water Supplies of Varying Reliability Served by Wells with Various Degrees of Coliform Contamination

Coliform Counts	Degree of Reliability			Number and Percent of Households
	Good	Medium	Poor	
Households with one or fewer coliforms per 100 milliliters of water	64.1	51.1	51.3	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	35.9	48.8	48.7	3,765,000 38.2
2 - 99 coliforms	21.8	28.2	28.7	2,264,000 23.0
100 + coliforms	14.1	20.6	20.0	1,501,000 15.2
Number of Households	8,057,000	1,385,000	4,059,000	9,850,000
Percent of Total	81.8	14.1	4.1	100.0

Table X-141

Percent of Households With Various Numbers of Water Treatment Devices Served by Wells with Various Degrees of Coliform Contamination

Coliform Counts	Number of Devices					Number and Percent of Households
	0	1	2	3	4+	
Households with one or fewer coliforms per 100 milliliters of water	47.5	59.5	61.9	67.4	58.4	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	52.5	40.5	38.1	32.6	41.6	3,765,000 38.2
2 - 99 coliforms	34.0	23.9	21.9	23.2	23.9	2,264,000 23.0
100 + coliforms	18.6	16.6	16.2	9.5	17.8	1,501,000 15.2
Number of Households	533,000	359,000	6,467,000	1,819,000	672,000	9,850,000
Percent of Total	5.4	3.6	65.6	18.5	6.8	100.0

problems declined—at least up to a point. Households served by wells which had four or more treatment devices had a higher rate of coliform problems than those with fewer devices. Missing to this analysis, of course, was the nature of the source water. It could have been that wells with numerous treatment devices started out with poorer quality water than other wells.

Connections

The wells under consideration are noncommunity system wells—that is, they had fewer than fifteen connections. Table X-142 divides wells into two categories: those which had one connection, and those with two to fourteen connections. About 80 percent of noncommunity system wells were individual household systems. The overall rates of coliform problems were roughly the same between the two groups of wells, although households served by two to fourteen connection wells had a slightly lower rate of problems. But, the rate of households with severe problems (total coliform counts of 100 or more per 100 milliliters of water) was almost twice as high (16.8 percent) among the one connection category as the two to fourteen connection category (8.8 percent).

Water test history

About 60 percent of all households served by wells reported that the water had never been tested to their knowledge. Table X-143 indicates that households with wells which had tested the water were slightly less likely to exceed the NSA coliform reference value. Systems of this small size very rarely tested with any regularity. Among the households which did test, most had done so only once, and the test results often could not be recalled. The tenuous history of testing, coupled with the total coliform results, suggested that what little testing did occur did not apparently result in alterations which affected the system's susceptibility to bacteria intrusion.

Table X-142

Percent of Households Served by Individual or Intermediate Wells with Various Degrees of Coliform Contamination

Coliform Counts	System Size Individual (one connection)	 Intermediate (2 - 14 connections)	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	61.4	63.1	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	38.5	36.9	3,765,000 38.2
2 - 99 coliforms	21.7	28.1	2,264,000 23.0
100 + coliforms	16.8	8.8	1,501,000 15.2
Number of Households	7,893,000	1,957,000	9,850,000
Percent of Total	80.1	19.9	100.0

Table X-143

Percent of Households Served by Wells with Various Levels of Coliform Contamination by Whether or Not the Household has Tested the Water

Coliform Counts	Test History Never Tested	 Tested at Some Time	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	60.3	64.0	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	39.7	36.0	3,765,000 38.2
2 - 99 coliforms	24.3	21.1	2,264,000 23.0
100 + coliforms	15.4	14.9	1,501,000 15.2
Number of Households	5,892,000	3,958,000	9,850,000
Percent of Total	59.8	40.2	100.0

Service years

Table X-144 shows the same four categories of years the household has been served by the major household water supply as was shown in Table X-129 for the nation as a whole. Other than a shift toward higher rates of coliform problems among wells, the two patterns were quite similar. Households served by wells from five to nine years had higher rates of coliform problems than those under five years. But, the ten to nineteen years of service group had a lower problem rate—indeed the lowest of all four categories. The longest length of service group, twenty years or more, again had higher rates.

As was pointed out in the discussion of Table X-129, the length of residence could not be precisely separated from the age of the system. Some households represented in the two lower categories, 0 to 4 and 5 to 9 years, were served by wells older than the inhabitant's length of residence and should have been assigned to the higher categories. Unfortunately, the necessary information to make a crisp distinction was unavailable.

Well type

The NSA classified the method of well construction into five categories, shown in Table X-145, along with a residual "other" category. Households served by drilled wells were by far the most common (67.1 percent). Households served by dug wells and bored wells demonstrated a coliform problem in more than half of the cases. Households served by dug wells had more than one coliform bacteria per 100 milliliters of water in over 76 percent of the cases. Almost half of the households using dug wells tested out with 100 or more coliforms per 100 milliliters. No other type of well construction approached the poor showing of dug wells.

Results for both driven and jetted wells were slightly more favorable than for drilled wells. But, both methods are restricted to favorable geologic

Table X-144

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Number of Years the System has Served the Household

Coliform Counts	Service Years 0 - 4	 5 - 9	 10 - 19	 20 +	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	63.7	55.1	67.8	58.2	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	36.3	44.9	32.1	41.8	3,765,000 38.2
2 - 99 coliforms	24.8	27.5	18.4	19.5	2,264,000 23.0
100 + coliforms	11.6	17.4	13.7	22.3	1,501,000 15.2
Number of Households	4,086,000	1,873,000	1,921,000	1,969,000	9,850,000
Percent of Total	41.5	19.0	19.5	20.0	100.0

Table X-145

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Mode of Well Construction

Coliform Counts	Well Type						Number and Percent of Households
	Dug	Bored	Drilled	Driven	Jetted	Other	
Households with one or fewer coliforms per 100 milliliters of water	23.6	49.8	68.1	72.6	70.9	53.3	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	76.4	50.3	31.9	27.4	29.1	46.6	3,765,000 38.2
2 - 99 coliforms	27.9	29.5	22.0	20.1	8.9	39.2	2,264,000 23.0
100 + coliforms	48.5	20.8	9.9	7.3	20.2	7.4	1,501,000 15.2
Number of Households	1,247,000	608,000	6,611,000	1,152,000	133,000	99,000	9,850,000
Percent of Total	12.7	6.2	67.1	11.7	1.4	1.0	100.0

conditions, such as sandy soil with few large rocks, conditions which can also provide some natural protection from surface water contamination. Though households served by jetted wells were slightly less likely than those using drilled wells to have coliform counts, it was the case that when the contamination did occur it was usually severe.

Grouting

Three categories were used to classify wells regarding grouting: grouted, not grouted, and unclassified (see Table X-146). The unclassified group consisted mainly of dug wells, where it made little sense to consider grouting as a useful technique for protecting from surface water intrusion around the well's exterior. As with dug wells (Table X-145), about three-quarters of the unclassified cases (Table X-146) had coliform counts higher than the NSA reference value.

For the remaining households, whether or not the well was grouted did not alter the rates of total coliform contamination. This finding did not suggest that grouting was a useless practice. It did not account for how well the grouting was done. It did not account for other aspects of the well construction which could inhibit intrusion. Some wells did not need grouting because of their depth or the nature of the soil they penetrated. Construction technique can obviate the need for grouting. For instance, cable tool wells have such small anulus spaces that grouting is unnecessary. Similarly, driven wells generally leave no space to accommodate grout. Thus, knowledge of whether or not a well was grouted was not sufficient alone to suggest the well was either more or less likely to be coliform-contaminated.

Well exit

Households served by wells which employed a pitless adapter or a buried seal were about equally likely to experience coliform problems. In each situation,

Table X-146

Percent of Households Served by Wells With Various Levels of Coliform Contamination by Whether or Not the Well was Grouted

Coliform Counts	Grouted	Not Grouted	Unclassified (mostly dug wells)	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	66.6	67.8	25.2	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	33.3	32.2	74.8	3,765,000 38.2
2 - 99 coliforms	22.2	22.2	28.2	2,264,000 23.0
100 + coliforms	11.1	10.0	46.6	1,501,000 15.2
Number of Households	4,381,000	4,189,000	1,280,000	9,850,000
Percent of Total	44.5	42.5	13.0	100.0

about 31 percent of the households had total coliform counts above one per 100 milliliters of water (see Table X-147). Households on wells which terminated above ground were about 14 percent more likely to experience coliform problems. They were also more than twice as likely to have high counts than households using wells with pitless adapters or buried seals.

Splashproof

Interviewers categorized the covers of wells as splashproof or not. The protection of wells from surface water intrusion was considered to be important in the reduction of contamination potential. Table X-148 supported that expectation. Households served by wells which were judged not to have splashproof covers had more than one coliform per 100 milliliters of water in over half (57.1 percent) of the cases. By comparison, those served by wells judged to be splashproof were above the NSA reference value in 33.6 percent of the cases. Furthermore, the proportion of households with 100 or more coliforms per 100 milliliters of water among the splashproof group was less than half that than those where the wells were judged not to be splashproof.

Cover

In a somewhat similar fashion, interviewers judged the general adequacy of the cover, beyond its ability to guard against surface water intrusion. Covers were generally categorized as adequate or not. Inadequate covers could be cracked, ill fitting, unsecured, and so on. The pattern was essentially identical to the judgments shown for whether or not the cover was splashproof. Covers not judged splashproof were usually judged inadequate regarding other aspects as well. Households served by wells with adequate covers were substantially less likely to demonstrate the presence of coliform bacteria (see Table X-149).

Table X-147

Percent of Households Served by Wells with Various Levels of Coliform Contamination According to How the Exit Pipe Leaves the Well

Coliform Counts	Pitless Adapter	Buried Seal	Neither (terminates above ground)	Unclassified	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	69.2	68.9	55.0	65.2	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	30.8	31.1	45.0	34.8	3,765,000 38.2
2 - 99 coliforms	21.9	21.4	24.0	29.3	2,264,000 23.0
100 + coliforms	8.9	9.7	21.0	5.5	1,501,000 15.2
Number of Households	2,842,000	1,803,000	5,072,000	134,000	9,850,000
Percent of Total	28.8	18.3	51.5	1.4	100.0

Table X-148

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Presence of a Well Cover which was Adjudged to be Splashproof or Not

Coliform Counts	Splashproof	Not Splash-proof	Unclassified	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	66.4	42.9	64.1	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	33.6	57.1	35.9	3,765,000 38.2
2 - 99 coliforms	21.0	30.5	30.9	2,264,000 23.0
100 + coliforms	12.6	26.6	5.0	1,501,000 15.2
Number of Households	7,769,000	1,935,000	146,000	9,850,000
Percent of Total	78.9	19.6	1.5	100.0

Table X-149

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Judged Adequacy of the Well Cover

Coliform Counts	Adequacy of Well Cover			Number and Percent of Households
	Adequate	Inadequate	Unclassified	
Households with one or fewer coliforms per 100 milliliters of water	66.3	45.6	63.1	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	33.7	54.4	37.0	3,765,000 38.2
2 - 99 coliforms	21.6	27.7	30.1	2,264,000 23.0
100 + coliforms	12.1	26.7	6.9	1,501,000 15.2
Number of Households	7,596,000	2,147,000	107,000	9,850,000
Percent of Total	77.1	21.8	1.1	100.0

Maintenance

Interviewers classified wells according to the general appearance and grouped them according to those which appeared to be adequately maintained and those which did not. Similar to the patterns for the splashproofness and general adequacy of the cover, Table X-150 shows that households using adequately maintained wells were less likely to exceed the NSA coliform reference value. The reference value was exceeded almost 15 percent more often among those households with wells judged to be inadequately maintained. Additionally, the inadequate maintenance group was more likely to be in the 100-plus coliform count group.

Depth

Table X-151 shows there was a dramatic reduction of coliform likelihood from households on very shallow wells (3 meters or less) to those which were slightly deeper (up to 10 meters). Households served by the shallowest wells were over the NSA reference value in 64.0 percent of the cases. There did not appear to be an important further reduction in risk until the depth of the well exceeded 20 meters. It should be noted that over 20 percent of all households served by wells of less than 20 meters depth had coliform counts of 100 or more coliforms per 100 milliliters. Households on wells that were from 30 to 60 meters in depth had the best showing. They were also the largest group. Interestingly, households on deeper wells of 60 meters and more were more likely to exceed the NSA reference point than the 30 to 60 meter group. That increased coliform likelihood centered in the 2 through 99 coliform count region.

Age

The age of the well did not seem to be related to the degree of coliform contamination. Households using wells in each of four age groups (Table X-152)

Table X-150

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Judged Adequacy of the Well Maintenance

Coliform Counts	Well Maintenance		Number and Percent of Households
	Adequate	Inadequate	
Households with one or fewer coliforms per 100 milliliters of water	64.7	50.3	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	35.3	49.6	3,765,000 38.2
2 - 99 coliforms	22.3	25.7	2,264,000 23.0
100 + coliforms	13.0	23.9	1,501,000 15.2
Number of Households	1,992,000	7,858,000	9,850,000
Percent of Total	20.2	79.8	100.0

Table X-151

Percent of Households Served by Wells with Various Degrees of Coliform Contamination by Categories of Well Depth

	Well Depth (in meters)							
Coliform Counts	3 meters or less	3.1-9.9 meters	10-14.9 meters	15-19.9 meters	20-29.9 meters	30-59.9 meters	60+ meters	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	36.0	51.2	55.5	55.3	64.2	72.6	63.0	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	64.0	48.8	44.5	44.7	35.8	27.4	37.0	3,765,000 38.2
2 - 99 coliforms	37.2	22.8	22.4	23.0	23.2	19.4	27.3	2,264,000 23.0
100 + coliforms	26.8	26.0	22.1	21.7	12.6	8.0	9.7	1,501,000 15.2
Number of Households	287,000	1,562,000	853,000	1,022,000	2,029,000	2,698,000	1,398,000	9,850,000
Percent of Total	2.9	15.9	8.7	10.4	20.6	27.4	14.2	100.0

Table X-152

Percent of Households Served by Wells with Various Degrees of Coliform Contamination by Categories of Well Age

Coliform Counts	Well Age				Number and Percent of Households
	5 Years or Less	5.1 - 10 Years	10.1 - 20 Years	20+ Years	
Households with one or fewer coliforms per 100 milliliters of water	61.6	63.9	61.2	60.4	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	38.3	36.2	38.8	39.6	3,765,000 38.2
2 - 99 coliforms	24.3	22.1	22.6	21.5	2,264,000 23.0
100 + coliforms	14.0	14.1	16.2	18.1	1,501,000 15.2
Number of Households	3,951,000	1,990,000	2,106,000	1,802,000	9,850,000
Percent of Total	40.1	20.2	21.4	18.3	100.0

had very similar rates of coliform contamination. There did appear to be a slight tendency for the coliform counts in the higher, 100 and above count category, among older wells.

Contamination distance

Interviewers were trained to identify potential contamination sources and measure their distance from the well. There was no way to judge the geologic potential of the contamination reaching the water source or any number of other important factors. It was reasoned that the likelihood of contamination should generally increase as the distance from the well head decreased. Table X-153 shows moderate support for that argument. There was a consistent, but weak, tendency for households on wells further from the closest identified potential contamination source to have lower rates of exceeding the NSA coliform reference value.

Wastewater disposal

Household wastewater can be a particularly rich source of bacteria and an important possible source of contamination to a well. Table X-154 categorizes wastewater disposal into three types. "Unconfined" referred to situations where wastewater was discharged to the open ground, open water (other than sites designed for such purpose), or privys. "Septic," of course, referred to the variety of septic type systems. "Public sewage" was the third category. There was a 30 percent difference in the rates of coliform contamination between the first and last categories. Households where the wastewater was unconfined had wells with coliform counts above the NSA reference value in 53.2 percent of the cases. On the other hand, more than three-quarters of the households using public sewers had wells which tested below the NSA coliform reference value. Households in the unconfined group were almost four times more likely to fall in the 100 or more

Table X-153

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Distance to the Closest Visible, Potential Contamination Source

Coliform Counts	Distance to Contamination				Number and Percent of Households
	0 - 5 Meters	6 - 10 Meters	11 - 15 Meters	Over 15 Meters	
Households with one or fewer coliforms per 100 milliliters of water	58.4	57.7	62.6	63.0	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	41.6	42.3	37.4	37.0	3,765,000 38.2
2 - 99 coliforms	23.6	20.6	26.2	22.5	2,264,000 23.0
100 + coliforms	18.0	21.7	11.2	14.5	1,501,000 15.2
Number of Households	1,311,000	1,012,000	1,388,000	6,139,000	9,850,000
Percent of Total	13.3	10.3	14.1	62.3	100.0

Table X-154

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Type of Wastewater Disposal Used at the Household

Coliform Counts	Unconfined	Septic	Public Sewer	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	46.8	62.2	76.1	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	.53.2	37.8	23.9	3,765,000 38.2
2 - 99 coliforms	30.2	22.7	17.8	2,264,000 23.0
100 + coliforms	23.0	15.1	6.1	1,501,000 15.2
Number of Households	582,000	8,939,000	330,000	9,850,000
Percent of Total	5.9	90.8	3.4	100.0

Table X-154

Percent of Households Served by Wells with Various Levels of Coliform Contamination by the Type of Wastewater Disposal Used at the Household

Coliform Counts	Unconfined	Septic	Public Sewer	Number and Percent of Households
Households with one or fewer coliforms per 100 milliliters of water	46.8	62.2	76.1	6,085,000 61.8
Households with more than one coliform per 100 milliliters of water	53.2	37.8	23.9	3,765,000 38.2
2 - 99 coliforms	30.2	22.7	17.8	2,264,000 23.0
100 + coliforms	23.0	15.1	6.1	1,501,000 15.2
Number of Households	582,000	8,939,000	330,000	9,850,000
Percent of Total	5.9	90.8	3.4	100.0

coliform category than those on public sewers. Households using septic systems were more than twice as likely to be in the high coliform group than those using public sewers. Most rural households used septic systems.

Discussion

The construction and maintenance characteristics of rural wells have been separately shown to be related to the incidence of coliform contamination. At a larger level, 76.3 percent of households served by wells which were over the NSA coliform reference value had at least one of the following characteristics: the well was dug or bored, the water line exited the well above ground level, the cover was not splashproof, the cover was otherwise inadequate, the general maintenance of the well was judged to be inadequate, or the well was 10 meters or less in depth. Table X-155 shows that households served by wells which had more of these characteristics were more likely to exceed the NSA total coliform reference value. Households served by wells in general were over the NSA reference value in 38.2 percent of all cases. Among those which had wells with none of the characteristics in Table X-155, 28.1 percent were over the reference value. But, households with wells which had at least four of the characteristics were over the reference value in more than two out of every three households.

Households where the combined household income was reported to be under $10,000, or where the head had less than a high school education, or where the dwelling unit was a mobile home, or where the occupants were renting, were also more likely to be above the NSA coliform reference value. The group of households which had at least one of those characteristics accounted for 70.7 percent of all the well-served households which were found above the coliform reference value. Table X-156 shows that the likelihood of a coliform problem climbed from 30.7 percent to 58.3 percent from the group of households with none of the characteristics to those with three or four of them.

Table X-155

Relationship Between the Count of Selected Well Characteristics* and the Probability of a Coliform Problem Among Rural Households Served by Wells

Count of Specified Well Characteristics*	Percent of Rural Households	Probability of a Coliform Problem (more (more than 1 coliform per 100 milliliters of water)
With 0 characteristics	31.4	28.1
With 1 characteristic	31.8	31.4
With 2 characteristics	13.4	37.2
With 3 characteristics	10.4	53.0
With 4 characteristics	6.8	66.2
With 5 characteristics	3.7	66.7
With 6 characteristics	2.5	77.4

N = 9,850,000
n = 1,315

*Well characteristics:
well was dug or bored;
water exits the casing above ground level;
well cover was not deemed splashproof;
well cover was judged to be generally inadequate protection against intrusion;
well was judged to be inadequately maintained;
well was ten meters or less in depth.

Table X-156

Relationship Between the Count of Selected Socioeconomic Characteristics* and the Probability of a Coliform Problem Among Rural Households Served by Wells

Count of Specified Well Characteristics	Percent of Rural Households	Probability of a Coliform Problem (more than 1 coliform per 100 milliliters of water)
With 0 characteristics	36.4	30.7
With 1 characteristic	34.4	40.1
With 2 characteristics	22.1	41.1
With 3 or 4 characteristics	7.1	58.3

N = 9,850,000
n = 1,315

*Socioeconomic characteristics:
combined household income less than $10,000 per year;
the head of household had less than a high school education;
dwelling unit was a mobile home;
residents were renting.

Table X-157 shows similar findings among households with selected supply and wastewater characteristics. The probability of a coliform problem ranged from 35.1 to 76.6 percent.

When the characteristics of all the last three tables were combined (Table X-158), the probability of being above the NSA coliform reference value ranged from 21.4 percent among households with none of these attributes to over 94 percent for those with ten or more.

Table X-157

Relationship Between the Count of Selected Water Supply and Wastewater Characteristics* and the Probability of a Coliform Problem Among Rural Households Served by Wells

Count of Specified Water Supply and Wastewater Characteristics	Percent of Rural Households	Probability of a Coliform Problem (more than 1 coliform per 100 milliliters of water)
With 0 characteristics	74.5	35.1
With 1 characteristic	20.3	44.1
With 2 characteristics	2.9	47.5
With 3 or 4 characteristics	2.2	76.6

N = 9,850,000
n = 1,315

*Water supply and wastewater characteristics:
- supply was judged to have poor accessibility;
- supply was judged to have poor to medium reliability;
- supply had no water treating devices;
- wastewater disposal was uncontrolled (privy, dump on ground, etc.).

Table X-158

Relationship Between the Count of Selected Well, Socioeconomic, and Supply and Wastewater Characteristics* and the Probability of a Coliform Problem Among Rural Households Served by Wells

Count of Specified Characteristics	Percent of Rural Households	Probability of a Coliform Problem (more than 1 coliform per 100 milliliters of water)
With 0 characteristics	12.4	21.4
With 1 characteristic	20.0	25.6
With 2 characteristics	19.3	36.5
With 3 characteristics	15.9	35.1
With 4 characteristics	11.7	43.2
With 5 characteristics	7.9	51.1
With 6 characteristics	5.9	64.7
With 7 characteristics	2.9	59.4
With 8 characteristics	1.9	73.5
With 9 characteristics	1.1	73.5
With 10 or more characteristics	1.1	94.3

*Well, socioeconomic, and supply and wastewater characteristics:

well was dug or bored;
water exits the casing above ground level;
well cover was not deemed splashproof;
well cover was judged to be generally inadequate protection against intrusion;
well was judged to be inadequately maintained;
well was ten meters or less in depth;

combined household income less than $10,000 per year;
the head of household had less than a high school education;
dwelling unit was a mobile home;
residents were renting;

supply was judged to have poor accessibility;
supply was judged to have poor to medium reliability;
supply had no water treating devices;
wastewater disposal was uncontrolled (privy, dump on ground, etc.).

REFERENCES

[1] Pipes, Wesley O. (editor). Water Quality and Health Significance of Bacterial Indicators of Pollution. Proceedings of a workshop held at Drexel University, Philadelphia, Pennsylvania. April 17-18, 1978.

[2] Sontheimer, Heinrich. "European Experience with Problems of Drinking Water Quality" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

[3] Rohlich, Gerard A. "Drinking Water and Health" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

[4] Atkinson, John W. "British Experiences on Problems of Drinking Water Quality" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

[5] Rohlich, 1978.

[6] Whyte, William Foote. Organizing for Agricultural Development. New Brunswick, NJ: Transaction Books, 1975.

[7] Sale, Kirkpatrick. Human Scale. New York: Coward, McCann and Geoghegan, 1980.

[8] Doerksen, Harvey R. and Pierce, John C., editors. "Citizen Influence in Water Policy Decisions: Context, Constraints and Alternatives" in Water Politics and Public Involvement. Ann Arbor: Ann Arbor Science Publishers, Inc., 1976.

[9] Baumann, Duane D. and Dworkin, Daniel M. "Consumer Perception and the Choice of Drinking Water Quality" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

[10] Baumann, 1978.

[11] Thomas, W. I. The Polish Peasant in Europe and America. New York: Dover Press, 1958.

References continued

[12] Andrews, Wade H.; Madsen, Gary E.; and Hardin, Clay W. Testing Social Indicators in the Techcom Model for Water Development. Logan: Utah State University, 1979.

[13] Gallup Organization, Inc. Public Attitudes Toward the Quality of Drinking Water. Princeton, NJ, 1974. Also, National Survey Conducted for the American Water Works Association. Princeton, NJ, 1973.

[14] Bruvold, William. "Human Perception and Evaluation of Water Quality," Critical Reviews in Environmental Control, Vol. 5, 1975, pp. 153-231.

[15] Saunders, Robert J. and Warford, Jeremy J. Village Water Supply: Economics and Policy in the Developing Third World. Baltimore: Johns Hopkins University Press, 1976.

XI

Factors Related to Rural Water Conditions: Quantity

Quantity was addressed in the NSA by examining demand for water at rural households. Demand was measured in terms of the amount of water households used in a month, calculated from water bills, and in terms of household respondents' perceptions of the amount usually provided by the household supply. These two measures are referred to as recorded quantity and perceived quantity, respectively. They are analyzed separately in the two major sections of this chapter.

The analysis presented in Chapter XI is an attempt to isolate factors which influence demand for water at rural households. Identification of such factors should be useful for planning the construction or modification of rural water supplies and for planning water conservation programs. Typically, demand for water is estimated using a "needs" approach based on growth in population and in per capita water use. However, incorporating the effects of water pricing and other factors would presumably improve the accuracy of demand estimates. Recognizing the impact of pricing on water use, some researchers have proposed various pricing strategies to promote water conservation.[1,2] One proposal, a summertime surcharge aimed at reducing lawn sprinkling, assumes that outdoor water use contributes significantly to total household demand for water. Though some empirical evidence exists to support this assumption, the dimensions of the relationship need to be specified so that the potential impact of pricing on household water consumption can be assessed more accurately.

Besides being relevant to the demand for water and to possible conservation practices, the quantity analysis presented in this chapter is aimed at identifying situations where water supply conditions are likely to be unsatisfactory with regard to quantity. At the extreme, a supply that was unsatisfactory with regard to quantity would be one that did not provide enough water to meet a household's basic sanitary needs.

The quantity information used in the analysis is presented in detail in Chapter V of this report. As explained there, information on recorded quantity was collected from household water bills. Meter readings provided the basic data on residential demand. Water bills covered periods ranging from a single month to an entire year, but the data were standardized to reflect monthly residential water consumption, reported in liters. Data were recorded from only one water bill per household, so that seasonal variation could not be assessed or taken into account. Most household water bills covered a period of one month, which may have resulted in an overestimation of demand, since most NSA interviews were conducted during summer months when water consumption was generally at a peak.

Also, though 9.2 million rural households received regular water bills, bills were not always available at the time of the NSA interview. Sometimes, too, a bill was available but lacked major components of information important to the NSA analysis—the amount charged, for instance, or the quantity consumed. When certain components were missing, the data were estimated from existing information. In all, the data base comprised 567 sampled households, representing 5.4 million households across the US, or nearly 60 percent of all households which received regular water bills. These households were served exclusively by community water systems and, therefore, the conclusions drawn for recorded quantity apply only to households served by systems with fifteen or more connections.

In addition to compiling consumption data from water bills, the NSA obtained information from respondents at all 2,654 sampled households on the quantity of water received from their water supplies during the year preceding the NSA. This information —referred to as perceived quantity—reflected respondents' assessments of water supply conditions over an entire year rather than actual consumption during a particular month. Perceived quantity information included both the amount of water supplied and the regularity of the condition. (Perceived quantity is discussed more fully, and results given, in Chapter V.)

Thus, the NSA analysis included two dependent variables—monthly household water consumption as derived from water bills (recorded quantity), and respondents' information on the general condition of the water supply with regard to quantity (perceived quantity). Independent variables to be included in the analysis were identified on the basis of theoretical considerations and past studies by other researchers. These independent variables, factors which might account for variation in household consumption and the amount of water generally supplied, were grouped into three categories to facilitate discussion: household attributes, supply attributes, and system attributes.

The two dependent variables, recorded quantity and perceived quantity, are analyzed separately in the sections to follow. In each major section, the independent variables that were examined are set forth first. Then bivariate relationships are explored, from which variables are selected for inclusion in models. Intermediate models are developed for household attributes, supply attributes, and system attributes. Finally, a model synthesizing the intermediate models is presented and conclusions drawn.

ANALYSIS OF RECORDED QUANTITY

HOUSEHOLD ATTRIBUTES

Sociodemographic characteristics such as age, education, and work status have been suggested as related in various ways to patterns of household water consumption. In particular, Spaulding has suggested that demand for water is one way in which cultural values are expressed.[3] For example, life styles (which vary according to ethnicity, education, and socioeconomic status) may have an important bearing on how water is used in a household and on the quantity of water used. Similarly, age and age-related status with regard to child-rearing may have a strong impact on household demand for water. Likewise, income may determine in large measure household demand for water, especially with regard to outdoor uses. The affordability of the supply—the cost of water in relation to total household income—may serve to restrict water usage, depending on the proportion of the household budget which can be allocated to water. The relationship between price and demand for water is well-documented.[4,5] Moreover, where water bills serve as documentation of charges as a function of demand, users may have a direct incentive for limiting consumption.

Rural residents' perceptions of water supply conditions also may influence household consumption. For example, if household members believe that the quality of the water supply is poor, they may curtail certain uses. Likewise, if they think the cost is high or if they are dissatisfied with some aspects of the supply, they may reduce consumption. Many factors related to residents' attitudes about their water supplies were considered in the analysis.

Many prior studies have explored the relationship of household size to water consumption and found—as one would expect—a positive relationship.[6-10] Certain other relationships also were explored—the length of time residents had lived in the household and been served by the same water supply, and housing tenure, or whether the housing unit was owned or rented by the residents. Both factors could be related to residents' willingness to make adjustments in their demand for water and in their consideration of seeking alternate supplies.

Various characteristics of the housing structure itself may be related to household demand for water. Grima proposed that housing density (single-unit versus multiple-unit structures) may be an important factor in estimating demand.[11] Howe and Linaweaver hypothesized that the age of the housing unit modifies consumption by affecting rates of flow and leakage, assuming that the age of the structure accurately reflects plumbing conditions.[12] Batchelor maintained that a demand model requires the inclusion of household water-related technology and functional usage (indoor and outdoor uses).[13]

Household attributes studied in the NSA analysis of rural residential demand for water are listed in Table XI-1. These variables were described in the introduction to the analysis chapters, but some were treated slightly differently in the quantity analysis than was indicated there. UNIT COST was originally described as the cost per 1,000 gallons of water. For the analysis in Chapter XI, a logarithmic transformation was applied to this variable to linearize its relationship to the quantity consumed at the household. Similarly, the variable WORKING was reduced to two categories, employed and not employed, because of the reduced case base in the analysis of recorded quantity (n = 567). The employed group consisted of household heads who were working either full time or part time, while the group that was not employed comprised household heads who were keeping house, going to school, retired, or unemployed. Finally, for the analysis of recorded quantity, OUTDOOR USES was treated as a categorical variable, which was consistent with prior research efforts. Households were divided into two groups: those that used the water supply for domestic purposes only, and those that used the water supply for both domestic and outdoor purposes. Although indoor uses also could affect household water consumption, they showed little variability in NSA results and so were not explored in the analysis.

Table XI-1

Independent Variables Used in the Analysis of Recorded Quantity

Continuous Variables	Categorical Variables
Household Attributes	
AGE EDUCATION INCOME PEOPLE RESIDENCE UNITS YEARS SERVED AFFORDABILITY DEVICES DISLIKED LIKED PERCEIVED QUALITY PERCEIVED QUANTITY SATISFACTION UNIT COST	ETHNICITY (white, nonwhite) OUTDOOR USES (domestic, domestic and outdoor) PERCEIVED COST (inexpensive, reasonable, expensive, very expensive) TENURE (own, rent, other) WILLINGNESS TO PAY (less, same, more) WORKING (employed, not employed)
Supply Attributes	
ACCESSIBILITY AESTHETIC/ECONOMIC QUALITY RELIABILITY	None
System Attributes	
CONNECTIONS DEBT RETIREMENT LINES O & M OPERATORS REVENUE REVENUE/CXN SYSTEM AGE	OWNERSHIP (public, private individuals and groups, private firm) SOURCE (groundwater, surface water, purchased, mixed)

SUPPLY ATTRIBUTES

Various characteristics of a water supply besides its quality may have a bearing on how a household uses the supply. A tendency towards breakdowns, for instance, may result in a household using less water, either intentionally or unintentionally. Likewise, the accessibility of a supply may affect its use.

The supply attributes examined in the quantity analysis are listed in Table XI-1. Since no modifications specific to the quantity analysis were necessary, these variables are the same as presented in the introduction to the analysis chapters.

SYSTEM ATTRIBUTES

Certain system attributes also may influence household demand for water. Public systems often are subsidized from general municipal funds; private systems, unable to rely on such subsidies and needing to realize some profit, may charge more for water. The higher cost may serve to lessen some kinds of water uses, and hence decrease demand.

Water system attributes examined in the quantity analysis are listed in Table XI-1. They are described in the introduction to the analysis chapters, but one modification occurred for the quantity analysis: OWNERSHIP was reduced to three categories from four because, in the smaller case base of the quantity analysis, forms of ownership in the "other" category were not encountered.

BIVARIATE RELATIONSHIPS

Distribution of recorded quantity

Water consumption—as derived from metering information on water bills—varied tremendously among rural households. Monthly use ranged from 695 liters (183 gallons) through 153,684 liters (40,573 gallons). The mean was 24,688 liters (6,518 gallons), while the median was approximately 17,750 liters (4,686 gallons).

The standard deviation of the distribution was 22,678 liters, reflecting the wide variation in monthly usage.

Household attributes

Data pertaining to sociodemographic characteristics and household water consumption are presented in Tables XI-2 and XI-3. Table XI-2 presents the correlations found between quantity and the continuous variables which were examined. Table XI-3 presents summary statistics for categorical variables. Four attributes—the number of household occupants (PEOPLE), the cost per thousand gallons (UNIT COST), affordability (AFFORDABILITY), and the number of water-using fixtures and devices (DEVICES)—showed a systematic pattern of correlation (r) with the quantity of water consumed monthly. Each had a coefficient of determination (R^2, eta^2) of 0.03 or more which was used as a criterion of analytical significance and retention for further modeling.

As one might expect, there was a positive relationship between household size and water consumption—as the number of occupants increased, so did household water consumption. The strength of the relationship was moderate ($r = 0.32$). The cost of water had a negative impact on consumption. As shown in Table XI-2, the correlation (r) was -0.47, accounting for fully 22 percent of the variability (R^2) in consumption. This negative relationship between cost per unit and consumption is depicted in Figure XI-1, which is a bivariate plot of all points on the two variables. As suggested by the figure, water consumption tends to be high when unit costs are lower, but tends to decline as unit costs increase. Although the figure illustrates the relationship between cost and consumption, individual values of UNIT COST cannot be interpreted directly because they have been transformed by applying the natural logarithm. This is the reason that negative values appear on the figure's horizontal axis; negative values correspond to unit costs of less than one dollar per thousand gallons.

Table XI-2

Correlations Between Recorded Quantity and Continuous Household Variables

Variable	r	R^2
AGE	-0.17	0.03
EDUCATION	0.10	0.01
INCOME	0.15	0.02
AFFORDABILITY	0.25	0.06*
UNIT COST	-0.47	0.22
PERCEIVED QUALITY	-0.09	0.01
SATISFACTION	0.13	0.02
LIKED	**	**
DISLIKED	-0.06	0.00
YEARS SERVED	-0.14	0.02
RESIDENCE	-0.15	0.02
PEOPLE	0.32	0.10
UNITS	-0.03	0.00
DEVICES	0.20	0.04

N = 5,374,000
n = 567

*Underscoring denotes that the variable was retained for further analysis.

**No liked characteristics were specified.

Table XI-3

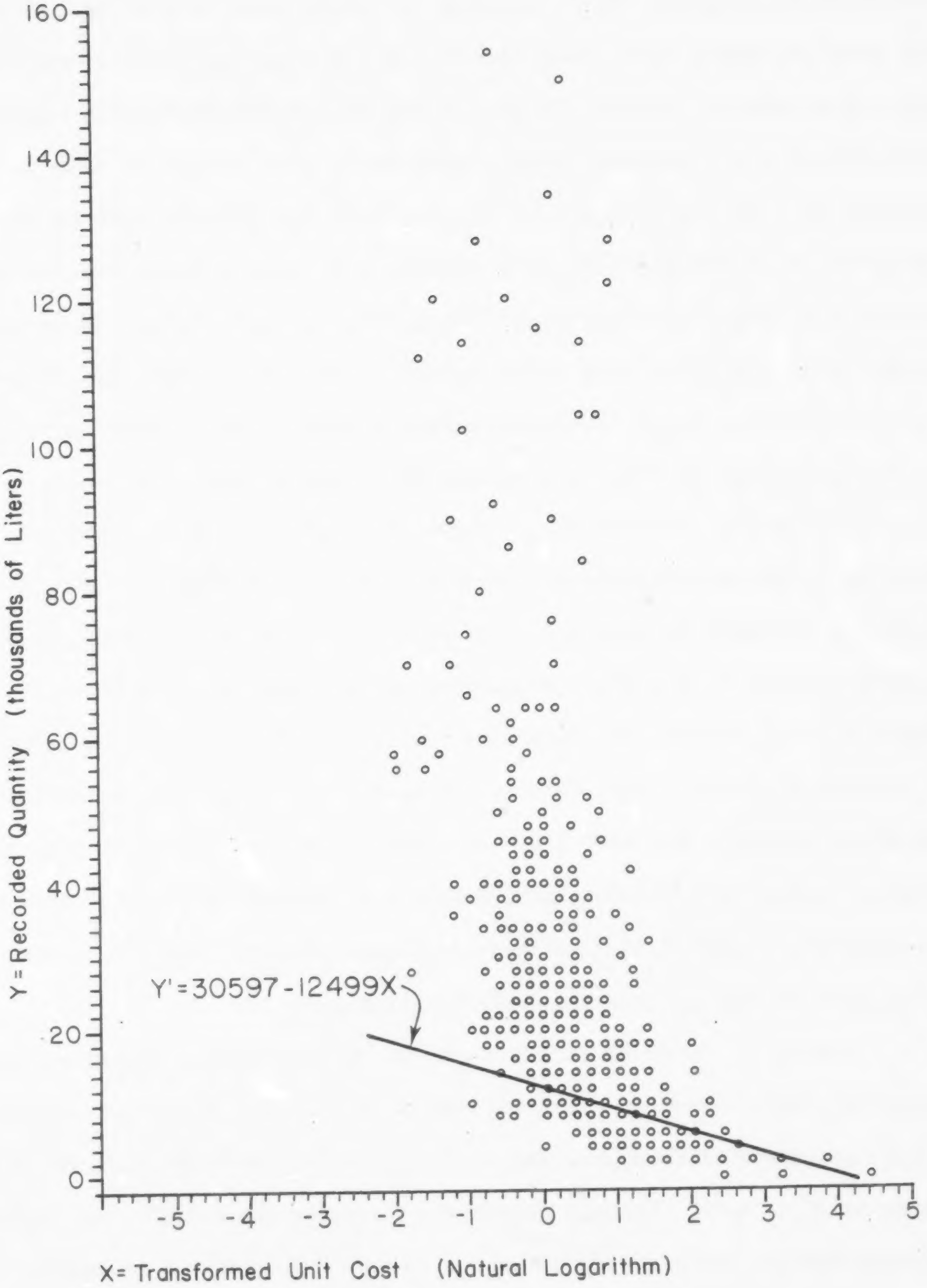
Summary Statistics for Recorded Quantity According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
ETHNICITY				0.00
White	88.8	24094	22019	
Nonwhite	11.2	28263	27917	
TENURE				0.00
Own	89.0	24434	22847	
Rent	10.2	25610	20666	
Other	0.8	24961	38654	
WORKING				0.02
Employed	69.0	26940	23058	
Not employed	31.0	19676	20959	
OUTDOOR USES				0.01
Domestic	6.0	15901	11256	
Domestic and outdoor	94.0	25249	23105	
PERCEIVED COST				0.03
Inexpensive	11.9	24427	18880	
Reasonable	65.5	23605	22220	
Expensive	15.8	22920	17179	
Very expensive	6.8	39698	35443	
WILLINGNESS TO PAY				0.01
Less	2.1	19174	13206	
Same	64.8	26110	23868	
More	33.1	22261	20399	

N = 5,374,000
n = 567

Figure XI-1

Bivariate Plot for Recorded Quantity and Unit Cost



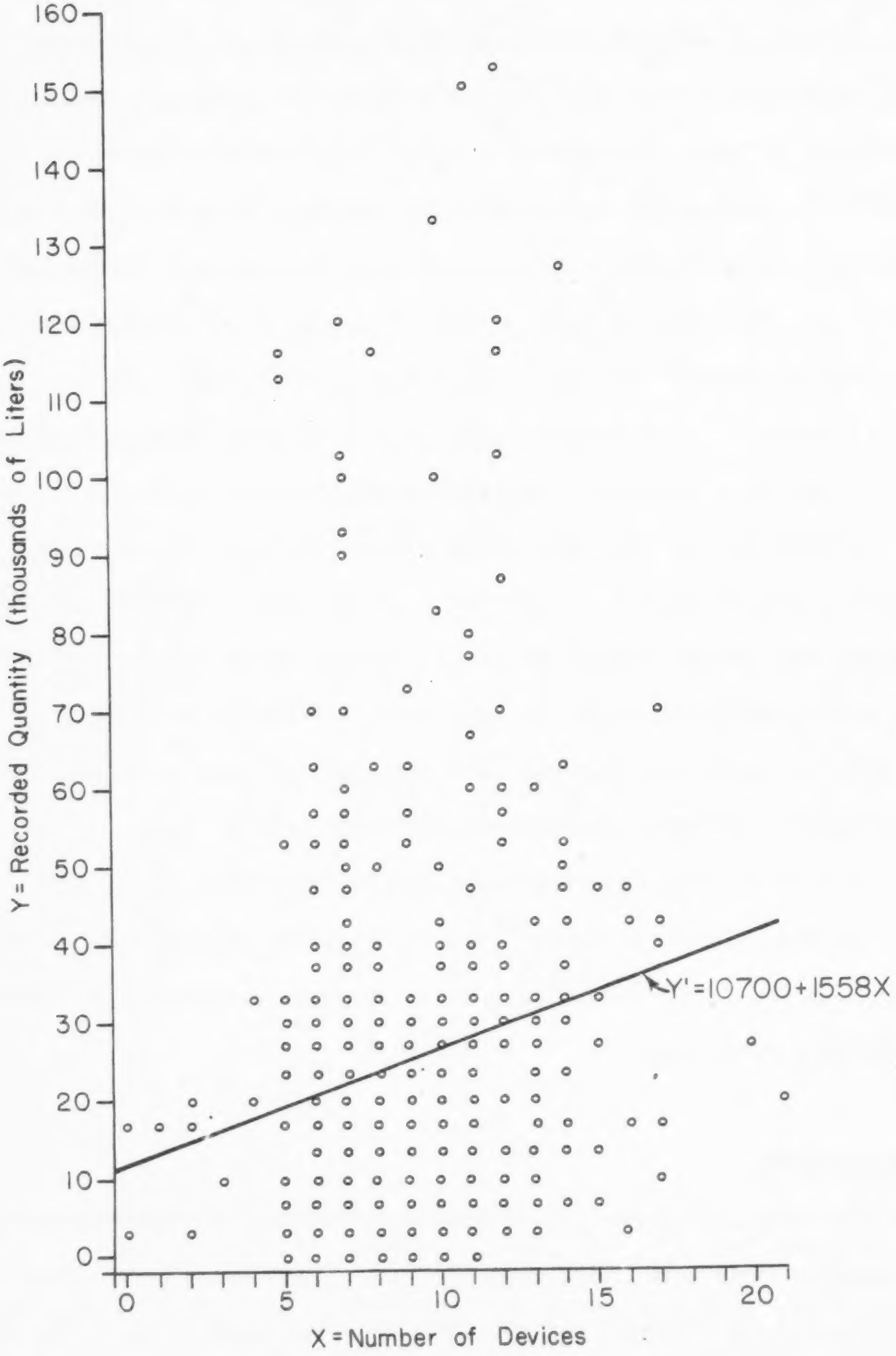
The finding that affordability was positively related to water consumption (r = 0.25) was somewhat unexpected, given the relationship between price (UNIT COST) and consumption. That is, one might intuitively expect that as the unit cost for water increased, water would become less affordable and demand would be reduced accordingly. Instead, the positive relationship with affordability suggests that families which allocated larger proportions of their budget for water were families that used more water. In Chapter V, it was reported that the mean proportion of income spent for water was nearly 1 percent, though the cost-to-income ratio ranged from 0.04 percent through 15.60 percent. Perhaps the cost-to-income ratio reflects the value placed on water, rather than a constraint imposed by a fixed budget. Stated differently, households which expend a larger proportion of their total budget for water may place a higher value on water as a commodity.

The number of water-using fixtures and devices at the household was a fairly good indicator of water consumption (r = 0.20). As illustrated in Figure XI-2, water use increased for each additional device up to the point of having twelve fixtures or devices; thereafter consumption declined with each additional item. There were two possible explanations for this pattern. First, dwellings which had an abundance of water-using fixtures and devices may once have served many household members who have since left home. Second, households with the greatest number of water-using fixtures may have required additional fixtures to meet household needs during peak demand periods (7 a.m. to 8 a.m., for instance), which were not used again for the remainder of the day.

Among status-related sociodemographic variables (that is, education, ethnicity, income, employment status, and housing tenure), only income and employment status (WORKING) approached the 0.03 criterion. In Table XI-2, INCOME shows an R^2 of 0.02. Although consumption averaged about 7,000 liters (1,848 gallons) more per month for households that had working heads than for households with heads who were not working (see Table XI-3), eta^2 for employment status fell

Figure XI-2

Bivariate Plot for Recorded Quantity and Number of Water-Using Fixtures and Devices



below 0.03, the criterion for selecting variables to include in the intermediate and final regression models.

Other household variables—including the age of the household head, the number of years in residence at the household, and the length of time served by the present water system—did not meet the criterion of significance established for the analysis. In Table XI-2, AGE shows an R^2 of 0.03, but this figure was rounded and did not actually satisfy the criterion. Interestingly, all three of these variables showed negative correlations with household water consumption. Perhaps length in residence and the number of years served by the water system simply reflected the age of the household head, which would show a relationship with consumption driven by family life cycle (in later stages, demand for water usually decreases).

Respondents' attitudes or perceptions about various factors associated with water supplies showed low correlations with the amount of water used. For instance, a greater general satisfaction with the supply (SATISFACTION) was correlated with greater demand for water. The correlation was low ($r = 0.13$), but it was in the expected direction. In Table XI-3, PERCEIVED COST shows an R^2 of 0.03, but this figure was rounded. The unrounded R^2 was below the criterion. Other attitudinal variables showed lower correlations with demand.

In a 1975 study, Batchelor found an association between demand and the variety of uses of the water supply,[14] but no association was apparent in the NSA analysis. Similarly, the number of units in the structure showed no systematic relationship with demand.

Supply attributes

The supply attributes did not show any important systematic relationships with variation in demand, as indicated by the low correlations presented in Table XI-4. Accessibility showed a positive relationship with consumption ($r = 0.12$), but reliability ($r = -0.01$) and the aesthetic/economic index ($r = -0.09$) were inversely

Table XI-4

Correlations Between Recorded Quantity and Continuous Supply Variables

Variable	r	R^2
RELIABILITY	-0.01	0.00
ACCESSIBILITY	0.12	0.01
AESTHETIC/ECONOMIC QUALITY	-0.09	0.01

N = 5,374,000
n = 567

correlated with consumption. The only unexpected finding was that supplies which were most accessible (as indicated by low scores on the accessibility index) tended to show lower demand.

System attributes

Like supply attributes, system attributes showed almost no systematic relationships with variation in demand (see Tables XI-5 and XI-6). Economic attributes showed very low correlations with demand; only system revenue (REVENUE) accounted for even 1 percent of the variability in consumption. One variable—the source of the system's water—accounted for 2 percent of the variability in household demand, but was below the criterion used in the analysis.

System ownership (OWNERSHIP) was the only variable among system attributes which was retained for further analysis. As shown in Table XI-6, this variable accounted for roughly 5 percent of the variability in household water consumption. Average monthly consumption (as indicated by the mean) was 37,454 liters (about 10,000 gallons) among households served by systems owned by private firms, compared to 22,711 liters (about 6,000 gallons) among households served by publicly owned systems.

INTERMEDIATE MODEL SPECIFICATION

The second stage of the analysis was to develop a separate regression model for each of the three groups of attributes in order to portray the overall impact of each set on the dependent variable, household demand for water. Each model included only those attributes which met the criterion of explaining at least 3 percent of the variability in demand ($R^2 \geq 0.03$) in the bivariate analysis. Among household attributes, four had met the criterion: the size of the household (PEOPLE), the number of water-using fixtures and devices at the household (DEVICES), the cost of the water supply (UNIT COST), and the affordability of the

Table XI-5

Correlations Between Recorded Quantity and Continuous System Variables

Variable	r	R^2
CONNECTIONS	0.05	0.00
LINES	0.04	0.00
SYSTEM AGE	-0.02	0.00
OPERATORS	0.05	0.00
REVENUE	0.09	0.01
O & M	0.04	0.00
DEBT RETIREMENT	-0.02	0.00
REVENUE/CXN	0.06	0.00

N = 5,374,000
n = 567

Table XI-6

Summary Statistics for Recorded Quantity According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SOURCE				0.02
Groundwater	49.9	27855	25511	
Surface water	22.3	19820	16098	
Purchased	10.9	21132	15387	
Mixed	16.9	24033	23581	
OWNERSHIP				0.05*
Public	85.5	22711	20077	
Private firm	9.9	37454	29341	
Private groups and individuals	4.6	33905	37595	

N = 5,374,000
n = 567

*Underscoring denotes that the variable was retained for further analysis.

water supply (AFFORDABILITY). No supply attributes met the analysis criterion. One system attribute met the criterion—the type of ownership of the system (OWNERSHIP).

Intermediate household model

The first step in the multivariate analysis was to assess the multicollinearity of the items to be considered. Table XI-7 is a matrix of intercorrelations of the selected household attributes. Zero-order correlations among the items were very low, indicating that collinearity was not a problem. Consequently, all selected variables were retained for the intermediate model.

The results of the regression analysis on household attributes, as well as the formal expression for the equation, are shown in Table XI-8. Interestingly, the model accounts for 45 percent of the variance in residential demand, about the same proportion as would be derived by simply summing the R^2s from Table XI-2 for all the variables represented in the model. In other words, each characteristic had a unique nonoverlapping association with residential water consumption.

As can be seen in Table XI-8, the unit cost of water had a dramatic effect on consumption. Specifically, for a unit increase in user charges (the natural log of dollars), household water consumption decreased by about 13,445 liters (about 3,550 gallons) per month.

In contrast to the negative effect of price on the demand for water, residential consumption increased in response to each additional person and water-using device at the household. Controlling for the other variables in the model, as the size of the household increased by one person, monthly water consumption increased by 3,349 liters (about 884 gallons). Similarly, an additional water-using fixture or device was associated with an increase in consumption of 1,434 liters (380 gallons).

Table XI-7

Intercorrelations Among Household Variables

	QUANTITY	AFFORD- ABILITY	UNIT COST	PEOPLE	DEVICES
QUANTITY	-				
AFFORDABILITY	0.250	-			
UNIT COST	-0.470	0.198	-		
PEOPLE	0.315	-0.041	-0.127	-	
DEVICES	0.207	-0.201	-0.107	0.175	-

Table XI-8

Intermediate Regression Estimates for Recorded Quantity and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	114.04
X_1	UNIT COST	b_1	-13445.43
X_2	PEOPLE	b_2	3349.06
X_3	AFFORDABILITY	b_3	6992.63
X_4	DEVICES	b_4	1434.23

$R^2 = 0.45$

Adjusted $R^2 = 0.45$

Standard error of estimate = 16,877.73

N = 5,374,000
n = 567

$$Y' = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Finally, with each additional 1 percent of total household income allotted for water, demand increased by 6,992 liters per month (about 1,846 gallons). In a sense, the affordability ratio represented a household's willingness to pay for water as a commodity of value.

Intermediate supply model

The bivariate analysis showed no systematic associations between residential demand for water and the supply attributes, therefore precluding intermediate model development.

Intermediate system model

Since only one system attribute was retained for inclusion in the intermediate model (OWNERSHIP), a regression estimate would merely duplicate the information from the bivariate analysis. Consequently, the information is not repeated here. Differences between each group mean and the unweighted average score on the dependent measure (the direct interpretation of the coefficients of categorically coded independent variables) can be computed from data presented earlier in the chapter (see Table XI-6).

FINAL MODEL SPECIFICATION

The final model of factors related to household demand for water incorporated system ownership and the household attributes which appeared in the intermediate model. Also, for reasons discussed in the preface to the analytical chapters, the design variables were also included in the final demand model.

In general, the final model yielded little improvement over the simpler intermediate household model. For instance, the increase in the amount of variance accounted for was only 3 percent (compare Tables XI-8 and XI-9), and the standard error of estimate was reduced by only 564 liters (a change of about 3

Table XI-9

Final Regression Estimates for Recorded Quantity

Variable Designation		Parameter	Estimate
	Intercept	a	9916.54
	REGION		
R_1	Northeast	b_1	-950.64
R_2	North Central	b_2	-1491.18
R_3	South	b_3	-981.78
	SMSA/NonSMSA		
S	SMSA	b_4	756.56
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	1184.17
SI_2	Small rural communities	b_6	-203.19
X_1	UNIT COST	b_7	-13910.89
X_2	PEOPLE	b_8	3082.24
X_3	AFFORDABILITY	b_9	6950.81
X_4	DEVICES	b_{10}	1269.88
	OWNERSHIP		
O_1	Private firm	b_{11}	3943.57
O_2	Public	b_{12}	-7875.14

$R^2 = 0.48$

Adjusted $R^2 = 0.48$

Standard error of estimate = 16,313.65

N = 5,374,000
n = 567

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7X_1 + b_8X_2 + b_9X_3 + b_{10}X_4 + b_{11}O_1 + b_{12}O_2$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

percent). These results implied that the design variables (region, SMSA/nonSMSA, and size of place) contributed very little to the predictive ability of the model. These results were not unexpected, however, since they corroborated earlier findings presented in Chapter V.

When direct comparisons were made of the intermediate and final models, the magnitudes of the regression coefficients were not significantly different. Thus, the interpretation is similar.

At a minimum, rural households required roughly 9,916 liters of water per month (2,618 gallons) to meet their needs (9,916 liters was the intercept). Based on estimates of family size, this quantity of water amounted to about 32 gallons per capita daily. Beyond the basic amount, however, water consumption varied directly with the size of the household, the number of water-using fixtures and devices at the household, and the proportion of total income allotted to water by the household. In contrast, the log of the price of water had a negative effect on demand. In fact, according to the regression estimates shown in Table XI-9, cost had the greatest impact of all variables on the amount of water consumed at the household. A unit increase in price (natural log of dollars) resulted in a decrease in demand of 13,911 liters (3,672 gallons) per month.

By comparison, as the size of the household increased by one person, average monthly demand increased by 3,082 liters (814 gallons). Likewise, the presence of one additional water-using fixture or device was accompanied by an increase in demand of 1,270 liters (335 gallons) per month. Finally, an increase from 1 to 2 percent of the total household income allocated to water was associated with an increase in demand of 6,951 liters (1,835 gallons) per month.

ANALYSIS OF PERCEIVED QUANTITY

The relationship between people's beliefs or perceptions about a particular situation and their behavior has long been a focus of social science research. In the previous chapter, this type of relationship was studied for water quality. Here, the relationship was explored in regard to the correspondence between people's assessments of the quantity of water readily provided by the household supply and the quantity of water actually used. However, NSA data allowed many other relationships to be explored as well.

The analysis presented in this section examines the influence of certain situational factors on perceptions about the quantity of water supplied to the household. Such factors included the type of supply serving the household, the ownership of the system, and the number of years the residents had lived in the household. The analysis does not go into detail about the association between specific NSA attitudinal data (such as the number of liked and disliked characteristics of the water supply) and the perceived quantity index, also an attitudinal measure. However, as might be expected, there was a close correspondence between the more specific perceptual data and the household's scores on the perceived quantity index, indicating a consistency in the responses of particular households.

As with the analysis of recorded quantity in the previous section, the analysis of perceived quantity begins with a listing of the independent variables that were examined. Bivariate analysis is used to select those variables which exhibited some systematic association with the index of perceived quantity. These variables are then included in intermediate models—one for household attributes, one for supply attributes, and one for system attributes. Finally, a model synthesizing the intermediate models is presented and conclusions drawn.

HOUSEHOLD ATTRIBUTES

Although there was a great deal of prior research which helped to identify factors associated with variation in recorded quantity, prior research was lacking which would be relevant to the analysis of perceived quantity. However, much of the reasoning applied in the analysis of recorded quantity also was applicable to perceived quantity.

The household attributes that were examined are listed in Table XI-10. They were presented in the introduction to the analysis chapters, but one variable (WORKING) was treated differently in this analysis from what was indicated there. Because of the uneven distribution of the cases, and consistent with the modification applied in the analysis of recorded quantity, categories for WORKING were reduced to two, employed and not employed. The employed group consisted of household heads who were working either full time or part time. The unemployed group consisted of household heads who were keeping house, going to school, retired, or unemployed.

SUPPLY ATTRIBUTES

Various attributes of household supplies were expected to influence people's perceptions of the quantity of water they usually received from their supplies. A number of supply attributes were examined, as listed in Table XI-10. The relationship between perceived quantity and the set of physical and chemical properties of supplies included in the index of AESTHETIC/ECONOMIC QUALITY was explored. Also examined was the influence of factors such as the type of supply, the reliability and accessibility of the supply, and certain characteristics of individual and intermediate wells.

SYSTEM ATTRIBUTES

Households served by intermediate and community systems often have no knowledge of various economic, physical, and operational characteristics of the system. The sole interaction with the water system may consist of paying a bill. Since households would have little knowledge of system characteristics, it was expected that there would be little association between perceived quantity and

Table XI-10

Independent Variables Used in the Analysis of Perceived Quantity

Continuous Variables	Categorical Variables
Household Attributes	
AGE EDUCATION INCOME PEOPLE RESIDENCE UNITS YEARS SERVED	ETHNICITY (white, nonwhite) TENURE (own, rent, other) WORKING (employed, not employed)
AFFORDABILITY DEVICES OUTDOOR USES UNIT COST QUANTITY	
Supply Attributes	
ACCESSIBILITY AESTHETIC/ECONOMIC QUALITY RELIABILITY WELL AGE WELL DEPTH	CONVEYANCE (pressurized public supply, electric pump, other) SUPPLY TYPE (community supply, well, other) WELL TYPE (drilled, driven, jetted, bored, dug)
System Attributes	
CONNECTIONS DEBT RETIREMENT LINES O & M OPERATORS REVENUE REVENUE/CXN SYSTEM AGE	OWNERSHIP (public, private individuals and groups, private firm, other) SOURCE (groundwater, surface water, purchased, mixed)

various system attributes. Nonetheless, various system attributes were examined, as listed in Table XI-10.

BIVARIATE RELATIONSHIPS

Distribution of scores on perceived quantity index

Scores on the perceived quantity index were based on the amount of water generally supplied to households, as reported by household respondents. All households in the NSA sample were asked for information on perceived quantity, whereas only households served by community systems supplied information for the analysis of recorded quantity.

Most rural households (80.7 percent) reported having ample water supplies. The remaining 19.3 percent of rural households reported various conditions during the previous year ranging from almost always receiving enough water from the supply to never receiving enough.

Because there was relatively little variation in the distribution of perceived quantity scores and because a disproportionately large number of households indicated having sufficient quantities of water throughout the year, it was decided to collapse the full range of categories on the index to two. (The index of perceived quantity consisted of ordered but discrete categories which could appropriately be collapsed.) The two new categories were designated Group 0 and Group 1 and contained 80.7 percent and 19.3 percent, respectively, of rural households. Group 0 included households where the water supply was reported to completely satisfy household quantity requirements; Group 1 included all households where the supply failed in some measure to provide enough water during the year preceding the NSA.

Household attributes

Relationships between perceived quantity and various household variables are displayed in Tables XI-11 and XI-12. As can be seen from the magnitude of the eta^2 coefficients, none of the household variables was systematically related to perceived quantity. No systematic relationship was evident between even the amount of water actually used at a household and perceived quantity. The lack of association between perceived and recorded quantity may have been induced by restrictions on the data, however. That is, this relationship could not be evaluated for households having individual water supplies, and some of those supplies might be expected to be inadequate in terms of quantity (hauled supplies and purchased bottled supplies, in particular).

Supply attributes

In general, water supply attributes seemed more related than household attributes to perceived quantity (see Tables XI-13 and XI-14). For instance, there seemed to be a slight association between supply breakdowns (RELIABILITY) and the amount of water a supply was perceived to generally provide. Group 0 households, which reported always having enough water, had an average score on RELIABILITY of 0.6, while the mean score for Group 1 households was 1.8. However, though the group means differed by a factor of three, RELIABILITY accounted for only 2.6 percent of the variation in perceived quantity. Since this fell slightly below the R^2 criterion of 0.03 (statistics reported in Table XI-13 are rounded), the variable RELIABILITY was not retained for further analysis.

Relationships between perceived quantity and various supply attributes also were explored for the subset of about 9.7 million households which were served by individual or intermediate wells (see Tables XI-13 and XI-14). One variable, WELL TYPE, showed a systematic relationship with perceived quantity. Specifically, the proportion of households in Group 1 having dug wells (0.36) was substantially

Table XI-11

Summary Statistics for Perceived Quantity According to Continuous Household Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
AGE	47.8	44.8	0.00
EDUCATION	11.6	12.3	0.00
INCOME	16509.7	18660.1	0.00
PEOPLE	3.0	3.2	0.00
RESIDENCE	10.4	10.3	0.00
UNITS	1.5	1.4	0.00
YEARS SERVED	9.2	10.5	0.00
AFFORDABILITY	3.5	3.1	0.00
DEVICES	8.4	8.9	0.00
OUTDOOR USES	1.7	1.8	0.00
N = 21,974,000 n = 2,654			
UNIT COST	2.6	2.1	0.00
QUANTITY	25289.1	23821.7	0.00
N = 5,374,000 n = 567			

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

Table XI-12

Summary Statistics for Perceived Quantity According to Categorical Household Variables

Variable and Categories	Actual Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
ETHNICITY				0.00
White	91.4	.81	.19	
Nonwhite	8.6	.76	.24	
TENURE				0.00
Own	79.4	.80	.20	
Rent	17.8	.81	.19	
Other	2.8	.88	.12	
WORKING				0.00
Employed	70.5	.79	.21	
Not employed	29.5	.84	.16	

N = 21,974,000
n = 2,654

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

Table XI-13

Summary Statistics for Perceived Quantity According to Continuous Supply Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
ACCESSIBILITY	7.5	7.7	0.00
AESTHETIC/ECONOMIC QUALITY	2.1	2.3	0.00
RELIABILITY	0.6	1.8	0.03
N = 21,974,000 n = 2,654			
WELL AGE	12.4	14.2	0.00
WELL DEPTH	37.8	28.2	0.01
N = 9,751,000 n = 1,300			

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

Table XI-14

Summary Statistics for Perceived Quantity According to Categorical Supply Variables

Variable and Categories	Actual Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
CONVEYANCE				0.00
Pressurized public supply	47.9	.82	.18	
Electric pump	46.9	.80	.20	
Other	5.2	.76	.24	
SUPPLY TYPE				0.01
Community supply	50.0	.82	.18	
Well	44.8	.82	.18	
Other	5.2	.64	.36	

N = 21,974,000
n = 2,654

WELL TYPE				0.03***
Drilled	67.8	.85	.15	
Driven	11.8	.83	.17	
Jetted	1.4	.91	.09	
Bored	6.2	.80	.20	
Dug	12.8	.64	.36	

N = 9,751,000
n = 1,300

---

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

***Underscoring denotes that the variable was retained for further analysis.

greater than the proportions of households in Group 1 with other types of wells ($\leq$0.20). Other supply attributes such as SUPPLY TYPE, CONVEYANCE, ACCESSIBILITY, and the AESTHETIC/ECONOMIC INDEX showed little association with perceived quantity, and none besides WELL TYPE was retained for further analysis.

System attributes

As expected, the system variables accounted for little variation in perceived quantity. In fact, only two variables, the number of system operators and the source of water, could account for even 1 percent of the variability in perceived quantity (see Tables XI-15 and XI-16).

INTERMEDIATE MODEL SPECIFICATION

Of the many household, supply, and system attributes that were examined for this analysis, only one variable, WELL TYPE, displayed a systematic relationship with perceived quantity that met the criterion for retention in subsequent analysis. Consequently, presentation of an intermediate model would only reiterate the relationship set forth in the bivariate stage. The analysis now proceeds to specification of the final model which incorporates the design variables and, for the subset of rural households served by wells, the type of well construction.

FINAL MODEL SPECIFICATION

The results of the regression model for perceived quantity are displayed in Table XI-17, along with the formal expression of the regression equation. As is evident from the table, the model accounted for almost none of the variability in perceived quantity. Furthermore, there was little difference in respondents' perceptions about the quantity of water they received from their supplies regardless of their location among the four regions, inside or outside SMSAs, or according to size of place.

Table XI-15

Summary Statistics for Perceived Quantity According to Continuous System Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
DEBT RETIREMENT	17.5	16.8	0.00
LINES	424.5	502.0	0.00
O & M	52.5	55.9	0.00
OPERATORS	12.5	17.5	0.01
REVENUE	1872102.0	2659041.0	0.00
REVENUE/CXN	360.5	659.6	0.00
SYSTEM AGE	16.7	19.0	0.00

N = 13,209,000
n = 1,461

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
CONNECTIONS	6276.2	8670.4	0.00

N = 21,974,000
n = 2,654

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

Table XI-16

Summary Statistics for Perceived Quantity According to Categorical System Variables

Variable and Categories	Actual Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
OWNERSHIP				0.00
Private groups and individuals	52.2	.80	.20	
Private firm	6.7	.85	.15	
Public	40.7	.81	.19	
Other	0.4	.71	.29	

N = 21,974,000
n = 2,654

Variable and Categories	Actual Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
SOURCE				0.01
Groundwater	58.8	.84	.16	
Surface water	18.1	.78	.22	
Purchased	9.5	.73	.27	
Mixed	13.6	.82	.18	

N = 13,209,000
n = 1,461

*Group 0 includes households which reported that the water supply completely satisfied household requirements throughout the year.

**Group 1 includes households which reported that the water supply did not satisfy all household requirements throughout the year.

Table XI-17

Final Regression Estimates for Perceived Quantity

Variable Designation		Parameter	Estimate
	Intercept	a	0.19
	REGION		
R_1	Northeast	b_1	-0.04
R_2	North Central	b_2	-0.02
R_3	South	b_3	0.00
	SMSA/NonSMSA		
S_1	SMSA	b_4	0.01
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.01
SI_2	Small rural communities	b_6	-0.01
	WELL TYPE		
WT_1	Dug	b_7	0.16
WT_2	Bored	b_8	-0.00
WT_3	Drilled	b_9	-0.04
WT_4	Driven	b_{10}	-0.02

$R^2 = 0.02$

Adjusted $R^2 = 0.02$

Standard error of estimate = 0.39

N = 9,751,000
n = 1,300

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7WT_1 + b_8WT_2 + b_9WT_3 + b_{10}WT_4$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

SUMMARY AND CONCLUSIONS

The purpose of the analyses presented in this chapter was to identify sets of factors which were systematically associated with variation in recorded quantity (monthly household water consumption) and perceived quantity (the respondent's perceptions of the amount of water generally received from the household supply). Among the many factors explored were household attributes, supply attributes, system attributes, and variables stemming from the design of the NSA sample. The results of the analysis of recorded quantity supported a standard postulate of economic theory—higher prices were associated with reduced consumption.

One limitation of the study was that the various pricing strategies actually used to derive user charges (for example, decreasing block and step rates, flat rate, uniform rate) could not be evaluated separately because of incomplete information. Instead, it was necessary to assume a uniform set of charges whereby water was priced at a constant rate for every thousand gallons consumed. Had the information on rate structures been available for those same households which provided sufficiently detailed cost and consumption information, different results may have been obtained. Nevertheless, the importance of pricing in moderating residential consumption was verified by the results of the analysis.

The analysis of recorded quantity identified a number of other important factors besides price. They were the average size of the household, the household's water-using fixtures and devices, and the ratio of water charges to total household income. As might be expected, consumption among rural households increased moderately with each additional occupant. Likewise, but to a lesser extent, consumption was affected by the number of water-using devices installed at the household. Finally, the cost-to-income ratio was found to be positively related to the quantity of water consumed at the household. This finding was somewhat unexpected, given the inverse relationship between price and household consumption. One possible explanation is that the cost-to-income ratio represented the consumer's willingness to pay for water, rather than a constraint imposed by a fixed budget. Further support for this idea is the fact that, compared to the price of other utilities or commodities purchased by rural households, water represents a very small proportion of the total household budget—usually 1 percent or less. Pricing trends indicate, however, that in coming years the cost of water is likely to increase as a proportion of the total household budget. As this occurs, the cost-to-income ratio may begin to affect water consumption patterns so they are more consistent with normal price-demand relationships.

One final point on the analysis of recorded quantity is noteworthy, though the results are difficult to interpret. Water consumption among households served by community systems showed great variation according to system ownership. Consumption was greatest when the system was owned by a private firm, somewhat lower when the system was owned by a private individual or group, and much lower when the system was owned by a public water company.

As to perceived quantity, the results of the regression analysis led to very limited conclusions. Most important was the finding that the regression model did not account for much of the variability in perceived quantity. The study identified only one variable—in effect, one situation—which seemed an important factor in accounting for insufficiencies in quantity. Specifically, dug wells were more often associated with insufficiencies than other types of well construction.

REFERENCES

[1] Hanke, Steve H. "Demand for Water Under Dynamic Conditions." Water Resources Research 6:5 (October 1970): pp. 1253-1261.

[2] Howe, Charles W., and Linaweaver, F. P. "The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure." Water Resources Research 3:1 (First Quarter 1967): pp.13-32.

[3] Spaulding, Irving A. "Social Class and Household Water Consumption." Social Behavior, Natural Resources, and the Environment. New York: Harper & Row, 1972.

[4] Howe and Linaweaver, 1967.

[5] Hanke, 1970.

[6] Gibbs, Kenneth C. "Price Variable in Residential Water Demand Model." Water Resources Research 14:1 (February 1978): pp. 15-18.

[7] Danielson, Leon E. Estimation of Residential Water Demand. Raleigh: North Carolina State University, 1977.

[8] Grunewald, O. C.; Haan, C. T.; Debertin, David L.; and Carey, D. I.; "Rural Residential Water Demand: An Econometric and Simulation Analysis." Water Resources Bulletin 12:5 (October 1976): pp. 951-961.

[9] Spaulding, 1972.

[10] Howe and Linaweaver, 1967.

[11] Grima, A. P. Residential Demand, Alternative Choices for Management. Toronto: University of Toronto Press, 1972.

[12] Howe and Linaweaver, 1967.

References continued

[13] Batchelor, R. A. "Household Technology and the Domestic Demand for Water." Land Economics 3 (August 1975): pp. 208-223.

[14] Ibid.

XII

Factors Related to Rural Water Conditions: Availability

Availability, as conceptualized in the NSA, refers to the ability of a water supply to provide water on a continuous basis. This definition of availability differs appreciably from the term's usual and more general conception, which emphasizes only the presence of water. The NSA assumed, for any particular household, that a water source existed and that there was some access to it. A second assumption was that the quantity of water that could be obtained from the source was sufficient to satisfy the household's demands under normal conditions. This combination of assumptions suggested that two specific aspects of the water supply were of primary importance in developing measures of availability. The first included the supply's accessibility, which was expressed as a combination of (a) the distance over which the water had to be transported, and (b) the pressure at which it was conveyed. A second aspect of availability consisted of the supply's reliability as reflected by the frequency of breakdowns and supply interruptions. Data on the components of these two variables were reported in Chapter V and summarized in composite form in Chapter VI. In this chapter, an effort will be made to establish the relationships between accessibility and reliability and selected household, supply, and system attributes.

The analytical strategy adopted for this chapter is similar to those employed in the chapters on quality (X) and quantity (XI). Initially, a series of bivariate models are developed to explore the systematic variation between a series of independent variables and measures of, first, accessibility, and second, reliability. The independent variables are organized according to whether they represent aspects of the household, aspects of the supply, or characteristics of the system, and are assessed individually as components of their respective groups. The purpose of this first stage of the analysis is to identify variables that appear to be associated with either reliability or accessibility. After variables from the three separate sets which are unrelated to the particular dependent variable under consideration are eliminated, intermediate models are constructed. Each intermediate model encompasses only those variables that were retained from the bivariate portion of the analysis. Finally, the variables selected for the intermediate models are combined into an extended or final model.

The dependent variables for this chapter are the composite measures of accessibility and reliability. The distributions of these two variables were described fully in Chapter VI, but relevant summary statistics are as follows. The mean score on the accessibility index was 7.5, the median was 7.6, and the standard deviation was 1.5. While the minimum score on the index was 0, the maximum was 47. The reliability index, in contrast, had a mean of 0.8, a median of 0.2, and a standard deviation of 2.8. Similar to the accessibility index, the minimum value on the reliability index was 0, while the maximum value was 52. In spite of these apparent similarities in the range of scores, the two indices were correlated at about -0.02. This independence reaffirmed the necessity for analyzing these variables separately and implied that their correlates would be unique.

Although they have not been specified in the literature on availability, certain relationships might be expected to emerge from this analysis. Some of these are intrinsic to the type of supply or system that provides water to the household, while others are inherent in the context of the household's water supply. Particular supply arrangements, for example, are designed to be more productive even with the same sources of water. Community water supplies that rely upon groundwater tend to have larger capacities than an individual household's well supply since the former distribute water to more connections and are more physically extensive, and because they support other functions besides domestic water consumption. The prominent additional function is, of course, fire protection, while large-scale commercial and industrial uses are also important. To produce the volume of water necessary for these functions, community supplies tend to transport water under higher pressures and through larger distribution lines. Additionally, since water must be constantly present, these supplies often have back-up facilities to avoid supply failures and interruptions or to minimize their effects. Redundancy—or at least complexity— in facilities may similarly limit the consequences of a breakdown to a portion of the supply connections rather than allowing the full complement of connections to be without water. Therefore, on the whole, community water supplies were anticipated to be more accessible and reliable than wells or other types of supply configurations.

If community supplies are distinguished by their dependability at the household level, cistern supplies represent the opposite condition. Since cisterns collect precipitation, they are more inclined than other supply types to be affected by the vagaries of weather. Also, because cisterns are exposed to the elements, their operation may be adversely influenced by debris and other material. The unreliability of cisterns limits their use except when other supply arrangements are not practicable. The particular characteristics of cisterns point up the probable relationship between supply type and reliability. In general, then, the type of supply was expected to show a significant relationship with both accessibility and reliability.

Besides detecting relationships that involve the various types of supplies, the following analysis may detect other associations. For instance, there may be a correspondence between reliability and the number of treatment and processing devices on the supply. This expectation is predicated on an assumption that the probability of supply disruptions increases with the number of devices attached to the supply because each one impinges independently on the supply's operation. Similarly, there may be a relationship between the quantity of water consumed at the household and the reliability of the supply. Increased water consumption may, to some extent, require more frequent use of equipment, which, in turn, may increase the likelihood of breakdowns. Since equipment tends to deteriorate over time, the age of the supply is another factor that may be implicated in the level of reliability. Likewise, the number of outdoor uses may be related to reliability because they could potentially exert unusual demands on the supply unless the supply was explicitly designed to accommodate them. The complete set of variables that may be related to accessibility and reliability will be presented in the next section of this chapter.

THE INDEPENDENT VARIABLES

The independent variables included in the analysis of the two dimensions of availability are provided in Table XII-1. This table arranges the variables according to whether they refer to household, supply, or system attributes and whether their form is continuous or categorical. Although the variables are essentially the same as those described in Chapter IX, there are a few differences that should be noted. More specifically, a number of the categorical variables were modified in the interests of facilitating the analysis of accessibility and reliability. For example, the type of housing tenure (TENURE), which initially consisted of three categories, now comprises only two (own and rent). Similarly, the work status of the household head (WORKING) encompasses only two

Table XII-1

Independent Variables Used in the Analysis of Availability

Continuous Variables	Categorical Variables
Household Attributes	
INCOME EDUCATION PEOPLE RESIDENCE YEARS SERVED UNITS LIKED DISLIKED DEVICES OUTDOOR USES QUANTITY	ETHNICITY (white, nonwhite) TENURE (own, rent) WORKING (employed, not employed) PREFERRED SYSTEM (different, present) PREFERRED SIZE (individual, intermediate, community) PREFERRED OWNER (public, private individuals and groups, private firm, other) WILLINGNESS TO PAY (less, same, more)
Supply Attributes	
AESTHETIC/ECONOMIC QUALITY TOTAL TREATMENT WELL DEPTH WELL AGE	SUPPLY TYPE (community supply, well, other) CONVEYANCE (pressurized public supply, electric pump, other) WELL TYPE (dug, bored, drilled, driven, jetted)
System Attributes	
CONNECTIONS OPERATORS SYSTEM AGE REVENUE REVENUE/CXN PERFORMANCE	OWNERSHIP (public, private individuals and groups, private firm, other) SOURCE (groundwater, surface water, other)

categories, working and not working, rather than the original five. The employed category was formed by collapsing two categories, working full time and working part time. The not employed category combines the categories of keeping house, retired, and unemployed. Except for these distinctions, the categorical variables are treated as set forth in Chapter IX. The continuous variables were not altered, so retain their earlier distributional properties.

ANALYSIS OF ACCESSIBILITY

It should be noted before specifying the bivariate models that the accessibility index was altered slightly from that presented in Chapter VI so that information on the entire complement of sampled rural households could be applied to the analysis. This modification involved arbitrarily assigning the maximum index value of 47 to those households that were formerly excluded from the index (that is, households that hauled water or carried it by hand). As a result of this change, the effective number of households in the sample was increased from 2,548 to 2,654, while the projected number increased from 21,290,000 to 21,974,000. The median score of this expanded accessibility index was 7.7, the mean value was 8.8, and the standard deviation was 7.0. Although this action certainly influenced the distribution of accessibility scores, it is probably still the case that a relative score of 47 for the 106 households underestimates the inaccessibility of these households because their supplies are the least accessible. That is, if index scores could have been independently compiled for these households, the scores most likely would have been higher than the arbitrary values that they actually received. The inclusion of these households was expected to improve the analysis by enhancing its accuracy.

BIVARIATE RELATIONSHIPS

The bivariate models are generally intended to establish the joint relationship between a single independent variable and a single dependent variable or, more typically, between any two variables. In this particular segment of the analysis, the independent variables are composed of household, supply, and system attributes, and the dependent variable is the accessibility index.

Household attributes

The results of the analysis of the relationships between the continuous household attributes and the accessibility index are reported in Table XII-2. Generally, the correlations do not depart dramatically from zero, which is the point at which no association exists, and are negative. The r's range in magnitude from 0.00 for the number of liked characteristics ascribed to the water supply (LIKED) to -0.32 for the number of water-using devices at the household (DEVICES). The largest positive correlation between the accessibility index and these continuous household attributes, which is 0.07, is reported for the number of years in residence (RESIDENCE). Assuming that they are accurate, these correlations demonstrate a moderate tendency for households with fewer water-using devices (DEVICES), lower education levels (EDUCATION), and a smaller number of outdoor uses (OUTDOOR USES) to have higher index scores and, therefore, less accessible water supplies. Similarly, the inverse relationship between INCOME and the accessibility index implies that households with lower incomes obtain water from supplies that are less accessible. Conversely, the positive association for length of residence (RESIDENCE) suggests that living in a household for a briefer period of time bears an association with greater supply accessibility. As their correlations indicate, accessibility is not systematically related to the other variables in the table.

Table XII-2

Correlations Between Accessibility and Continuous Household Variables

Variable	r	R^2
INCOME	-0.08	0.01
EDUCATION	-0.16	0.03*
PEOPLE	0.02	0.00
RESIDENCE	0.07	0.00
YEARS SERVED	0.02	0.00
UNITS	-0.05	0.00
LIKED	0.00	0.00
DISLIKED	-0.05	0.00
DEVICES	-0.32	0.10
OUTDOOR USES	-0.17	0.03
N = 21,974,000 n = 2,654		
QUANTITY	0.02	0.00
N = 5,696,000 n = 603		

*Underscoring denotes that the variable was retained for further analysis.

The general absence of appreciable relationships between the accessibility index and the continuous household attributes is manifested in the magnitude of the coefficients of determination (R^2) that are reported in Table XII-2. Seven of the eleven R^2s presented in the table are 0.00 and another is 0.01. However, two of the coefficients are 0.03 and the R^2 for DEVICES is 0.10. Since these three variables have coefficients at least equal to the criterion level of 0.03, they are candidates for the intermediate and final models.

Table XII-3 provides the relevant statistics on the relationships between the categorical household attributes and the accessibility index. As can be observed from the table, the means on the index for the categories do not differ substantially for most of the variables. There are two prominent exceptions: the preferred system (PREFERRED SYSTEM), where the mean accessibility score for households that prefer a different system (11.7) is much higher than the mean of the households that prefer their present system (8.1); and the preferred system ownership (PREFERRED OWNER), where the means for households preferring ownership by private firms (7.1) or public entities (7.7) are the smallest. These two categorical variables also generated the largest squared etas—0.04 for both sets of categories. The coefficients suggest that about 4 percent of the variation in supply accessibility may be accounted for by knowledge of the household's preferences for the system supplying its water and preferences for the form of the system's ownership. These two variables will be retained for subsequent parts of this analysis since their coefficients exceed the required inclusion criterion.

Supply attributes

The analysis of supply attributes is more consequential in terms of the magnitudes of the relationships that emerge. Table XII-4 reports the simple product-moment coefficients and R^2s produced by the bivariate models for selected supply characteristics and the accessibility index. The correlations tend

Table XII-3

Summary Statistics for Accessibility Index According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
ETHNICITY				0.01
White	91.4	8.5	6.2	
Nonwhite	8.6	11.4	12.1	
TENURE				0.00
Own	79.4	8.6	6.5	
Rent	20.6	9.5	8.6	
WORKING				0.01
Employed	70.5	8.4	5.9	
Not employed	29.5	9.6	9.0	
PREFERRED SYSTEM				0.04*
Different	17.8	11.7	12.2	
Present	82.2	8.1	5.0	
PREFERRED SIZE				0.02
Individual	39.6	10.0	8.8	
Intermediate	10.8	9.0	5.7	
Community	49.6	7.7	5.3	
PREFERRED OWNER				0.04
Public	41.0	7.7	5.2	
Private individuals and groups	51.7	9.6	8.0	
Private firm	6.3	7.1	1.3	
Other	1.0	11.2	11.2	
WILLINGNESS TO PAY				0.01
Less	3.6	9.5	8.7	
Same	63.3	8.2	5.2	
More	33.1	9.8	9.3	

N = 21,974,000
n = 2,654

*Underscoring denotes that the variable was retained for further analysis.

Table XII-4

Correlations Between Accessibility and Continuous Supply Variables

Variable	r	R^2
AESTHETIC/ECONOMIC QUALITY	0.03	0.00
TOTAL TREATMENT	-0.22	0.05*
N = 21,974,000 n = 2,654		
WELL DEPTH	-0.05	0.00
WELL AGE	-0.01	0.00
N = 9,751,000 n = 1,300		

*Underscoring denotes that the variable was retained for further analysis.

to be negative and range from a minimum of -0.01 for the age of the well (WELL AGE) to a maximum of -0.22 for the number of treatment devices (TOTAL TREATMENT). The R^2s, by comparison, are 0.00 except for the one involving TOTAL TREATMENT (0.05). Since the R^2 for TOTAL TREATMENT is above the stipulated value of 0.03, it is the only continuous variable retained for the multivariate supply models.

Although fewer variables, the categorical supply attributes demonstrated stronger patterns of relationship with accessibility. Table XII-5, which contains the appropriate statistics, indicates that the means by which water is conveyed to the household (CONVEYANCE) and the type of supply (SUPPLY TYPE) both show an association with scores on the accessibility index. The eta^2 for CONVEYANCE, the larger of the two coefficients, is 0.44, while that for SUPPLY TYPE is 0.27. In addition, the other eta^2 in the table, which is 0.03, suggests that type of well construction (WELL TYPE) and accessibility are systematically interrelated. While their coefficients differ, all three categorical supply attributes are influential enough to be included in subsequent models.

System attributes

A somewhat less distinctive pattern of interrelationships emerges from the analysis based on system attributes. The statistics showing associations between the relevant continuous system characteristics and the accessibility index are presented in Table XII-6. While three of the six variables in the table are positively correlated with accessibility, the other three have negative product-moment coefficients. Similarly, half the attributes have correlations equal to or greater than 0.10, although only one of these associations is positive. The maximum correlation is -0.11 for number of operators (OPERATORS) and the minimum is 0.03 for annual revenue per connection (REVENUE/CXN). None of the R^2s for the six continuous system variables meets or exceeds the 0.03 level.

Table XII-5

Summary Statistics for the Accessibility Index According to Categorical Supply Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SUPPLY TYPE				0.27*
Community supply	50.0	7.0	1.9	
Well	44.8	9.0	5.9	
Other	5.2	23.7	18.7	
CONVEYANCE				0.44
Pressurized public supply	47.9	6.9	1.0	
Electric pump	46.9	8.5	4.0	
Other	5.2	28.3	19.4	

N = 21,974,000
n = 2,654

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
WELL TYPE				0.03
Dug	12.8	11.6	11.1	
Bored	6.2	8.6	4.1	
Drilled	67.8	8.7	5.0	
Driven	11.8	8.1	1.8	
Jetted	1.4	8.0	1.6	

N = 9,751,000
n = 1,300

---

*Underscoring denotes that the variable was retained for further analysis.

Table XII-6

Correlations Between Accessibility and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.10	0.01
N = 21,974,000 n = 2,654		
OPERATORS	-0.11	0.01
SYSTEM AGE	0.10	0.01
REVENUE	0.08	0.01
REVENUE/CXN	0.03	0.00
PERFORMANCE	-0.09	0.01
N = 13,209,000 n = 1,461		

Consequently, the variables in the table are discarded from further consideration because they would make no appreciable contribution to the multivariate models.

As seen from Table XII-7, only one system variable is a candidate for the multivariate portions of the accessibility analysis. Specifically, the eta^2 for OWNERSHIP (0.05) suggests that a sufficient proportion of the variation in accessibility may be accounted for by variation on this attribute.

INTERMEDIATE MODEL SPECIFICATION

The preceding portion of this analysis assessed individual relationships between the accessibility index and three mutually exclusive sets of attributes. The following intermediate models attempt to show the interrelationships between the variables in each set and their combined association with accessibility. Each model contains a single dependent variable, the accessibility index, and two or more independent variables (those categorical or continuous household, supply, or system attributes retained from the bivariate analyses).

Intermediate household model

The intermediate model developed for the household attributes consists of the accessibility index and the five independent variables that were retained from the bivariate portion of the analysis. Three of these variables are continuous, while two are categorical. The categorical variables, PREFERRED SYSTEM and PREFERRED OWNER, are represented by sets of coded vectors containing information about the category or group membership. Consistent with the requirements of the regression framework, the number of coded vectors is equal to one less than the number of categories making up the variable. Therefore, although PREFERRED SYSTEM contains two categories, it designates only one coded vector in the model. Similarly, PREFERRED OWNER has three coded vectors even though it actually contains four categories. The equation for this

Table XII-7

Summary Statistics for the Accessibility Index According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
OWNERSHIP				0.05*
Public	40.7	7.2	3.4	
Private individuals and groups	52.3	10.1	8.8	
Private firm	6.6	7.5	3.8	
Other	0.4	17.0	16.8	
N = 21,974,000 n = 2,654				
SOURCE				0.02
Groundwater	58.8	7.6	2.8	
Surface water	18.1	7.0	2.5	
Other	23.1	6.8	1.1	
N = 13,209,000 n = 1,461				

*Underscoring denotes that the variable was retained for further analysis.

model is presented at the bottom of Table XII-8, which also reports the results of fitting the model to the data on accessibility and the five household attributes.

The regression coefficients in the table, which are interpreted differentially in this context, are negative and range from a maximum of -2.41 for the private firm category of PREFERRED OWNER to a minimum of -0.03 for EDUCATION. The R^2 for the intermediate household model, which is presented at the bottom of Table XII-8, indicates that 16 percent of the variation in accessibility can be accounted for by knowledge about rural households' water-using devices, outdoor uses of water, education levels, preferences for another water system, and preferences for system ownership. Although this result suggests that the model is not a particularly efficacious explanatory device, it is moderately interesting when compared to the coefficients produced by the bivariate models. These coefficients, which sum to 0.24, tend to differ appreciably from the R^2 of 0.16 that emerges when the variables are considered simultaneously. Since an R^2 of 0.16 is substantially lower than the 0.24 that would be expected if the attributes were independent, it is probable that some of them are redundant or convey similar information.

Intermediate supply model

The intermediate supply model, like the household model, is composed of a dependent variable and a set of independent variables. Again, these independent variables are a mix of continuous and categorical supply variables that were selected on the basis of their bivariate relationships with accessibility. The continuous attribute retained is the number of treatment devices connected to the supply (TOTAL TREATMENT), while the categorical variables are type of supply (SUPPLY TYPE) and the mode of conveyance (CONVEYANCE). For analytical purposes, both categorical attributes are represented by two coded vectors. The

Table XII-8

Intermediate Regression Estimates for the Accessibility Index and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	17.51
X_1	DEVICES	b_1	-0.61
X_2	OUTDOOR USES	b_2	-0.61
X_3	EDUCATION	b_3	-0.03
	PREFERRED SYSTEM		
PS_1	Different	b_4	-1.53
	PREFERRED OWNER		
PO_1	Public	b_5	-0.40
PO_2	Private individuals and groups	b_6	-2.05
PO_3	Private firm	b_7	-2.41

$R^2 = 0.16$

Adjusted $R^2 = 0.16$

Standard error of estimate= 6.39

N = 21,974,000
n = 2,654

$$Y' = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4PS_1 + b_5PO_1 + b_6PO_2 + b_7PO_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

expression for this model is included in the lower portion of Table XII-9 along with the parameter estimates and other relevant statistics.

While the intercept for the model is 16.88, the regression coefficients range from 0.18 for the well category of SUPPLY TYPE to -7.94 for the electric pump category of CONVEYANCE. These categorical coefficients represent mean differences on the accessibility index after the effect of TOTAL TREATMENT is partialed out of the relationship. Alternatively, the coefficient for TOTAL TREATMENT reflects the change in accessibility attributable to an incremental shift in this factor given the influence of the categorical variables.

The R^2 suggests that 52 percent of the variation in accessibility can be accounted for by the independent variables in the intermediate supply model. This represents a 36 percent improvement over the household model and an 8 percent increase over the degree of association seen for CONVEYANCE by itself. However impressive, this relative increase is not as large as it would be if the independent variables were mutually exclusive. Since households with community supplies also tend to be located in the pressurized public category, and households with wells are inclined to be in the electric pump category, this redundancy involves SUPPLY TYPE and CONVEYANCE.

To take maximum advantage of the bivariate results, another model is developed that includes only rural households that obtain water from well supplies. This model is essentially the same as the intermediate model just presented, but with a few prominent exceptions. First, and most important, this second intermediate supply model estimates relationships between variables for a distinct subset of households rather than for the entire sample. Second, although the models have similar structures, the subset model involves a slightly different configuration of independent variables. Type of supply (SUPPLY TYPE) is omitted from the model because it now is constituted by a single category, and type of well construction (WELL TYPE) is added as a component. Therefore, the subset model

Table XII-9

Intermediate Regression Estimates for the Accessibility Index and Supply Attributes[1]

Variable Designation		Parameter	Estimate
	Intercept	a	16.88
X_1	TOTAL TREATMENT	b_1	-0.37
	SUPPLY TYPE		
ST_1	Community supply	b_2	-7.04
ST_2	Well	b_3	0.18
	CONVEYANCE		
C_1	Pressurized public supply	b_4	-2.04
C_2	Electric pump	b_5	-7.94

$R^2 = 0.52$

Adjusted $R^2 = 0.52$

Standard error of estimate = 4.86

N = 21,974,000
n = 2,654

$$Y' = a + b_1X_1 + b_2ST_1 + b_3ST_2 + b_4C_1 + b_5C_2$$

Table XII-9 continued

Variable Designation		Parameter	Estimate
	Intercept	a	16.89
X_1	TOTAL TREATMENT	b_1	-0.81
	CONVEYANCE		
C_1	Pressurized public supply	b_2	-6.93
C_2	Electric pump	b_3	-7.27
	WELL TYPE		
WT_1	Dug	b_4	1.27
WT_2	Bored	b_5	0.12
WT_3	Drilled	b_6	0.31
WT_4	Driven	b_7	0.20

$R^2 = 0.54$

Adjusted $R^2 = 0.54$

Standard error of estimate = 4.03

N = 9,751,000
n = 1,300

$$Y' = a + b_1X_1 + b_2C_1 + b_3C_2 + b_4WT_1 + b_5WT_2 + b_6WT_3 + b_7WT_4$$

---

[1] The first model in the table includes all rural households, while the second applies only to households with well supplies.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

includes three variables—one continuous and two categorical. However, the number of parameters increases since WELL TYPE comprises four coded vectors instead of the two that represented SUPPLY TYPE. The formal expression for the subset intermediate supply model is presented in the bottom portion of Table XII-9, which also reports the results of fitting the model to the relevant data on supply attributes and accessibility.

The electric pump category of CONVEYANCE has the largest coefficient; the smallest coefficient is 0.12 for the bored category of WELL TYPE. The model's intercept of 16.89 is virtually the same as the one obtained for the full intermediate model. Similarly, the R^2 of 0.54 suggests that the subset model accounts for approximately the same proportion of the variance in accessibility as the model specified for all rural households. This also signifies that most of the systematic variation in accessibility could be accounted for by type of supply and that the addition of type of well construction (WELL TYPE) does not contribute appreciably to an understanding of accessibility. Consequently, type of well construction will be omitted from the final model even though it was significantly related to accessibility at the bivariate level.

Intermediate system model

Since only a single system variable was related to accessibility at the bivariate level, an intermediate system model incorporating form of ownership (OWNERSHIP) would merely duplicate the information from the bivariate analysis. Therefore, the information will not be repeated here, but the variable will appear in the specification of a final accessibility model.

FINAL MODEL SPECIFICATION

The final accessibility model combines the variables that appear in the preceding set of intermediate models and form of ownership into a single structure.

In addition to the various components of the intermediate models, the final model also includes the design variables. Twelve variables constitute the final accessibility model—four continuous and the remaining eight categorical. The formal expression for the final model appears in the bottom of Table XII-10, which also reports the estimates generated by fitting the final accessibility model to the relevant data.

According to the table, the intercept of the model is 19.39 and the magnitude of the largest coefficient is -7.61. Other regression coefficients tend to be small, with only four exceeding 1.0. In order of decreasing size, these correspond to the electric pump category of CONVEYANCE (-7.61), the community supply category of SUPPLY TYPE (-6.68), the pressurized public supply category of CONVEYANCE (-1.56), and the private firm category of PREFERRED OWNER (-1.12). While the coefficients of the other categorical variables are generally smaller than these four, those for the design effects (REGION, SMSA, and SIZE OF PLACE) are typically the smallest. When only continuous variables are considered, the coefficients for TOTAL TREATMENT (-0.35) and OUTDOOR USES (-0.33) are the largest. The R^2 for the model is 0.54.

The statistics presented in Table XII-10 suggest several conclusions about the final model developed to account for differences in accessibility. First, when the effects of other independent variables are held constant, the factors incorporated into the sample design are not important as sources of variation. Second, since its R^2 is less than anticipated, the final model does not improve the understanding of variation in accessibility beyond the level provided by the intermediate supply model. Third, this diminution in the final model's power is probably attributable to confounded variables that convey redundant information. This is partially confirmed by the material in Table XII-11, which summarizes the results for all three multivariate models. Comparisons across the final model and the appropriate intermediate model indicate that only the coefficients for TOTAL

Table XII-10
Final Regression Estimates for the Accessibility Index

Variable Designation		Parameter	Estimate
	Intercept	a	19.39
	REGION		
R_1	Northeast	b_1	-0.47
R_2	North Central	b_2	0.61
R_3	South	b_3	0.19
	SMSA/NonSMSA		
S_1	SMSA	b_4	0.15
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.29
SI_2	Small rural communities	b_6	0.00
X_1	DEVICES	b_7	-0.17
X_2	OUTDOOR USES	b_8	-0.33
X_3	EDUCATION	b_9	-0.01
	PREFERRED SYSTEM		
PS_1	Different	b_{10}	-0.71
	PREFERRED OWNER		
PO_1	Public	b_{11}	-0.17
PO_2	Private individuals and groups	b_{12}	-0.74
PO_3	Private firm	b_{13}	-1.12
X_4	TOTAL TREATMENT	b_{14}	-0.35
	SUPPLY TYPE		
ST_1	Community supply	b_{15}	-6.68
ST_2	Well	b_{16}	0.15

Table XII-10 continued

Variable Designation		Parameter	Estimate
	CONVEYANCE		
C_1	Pressurized public supply	b_{17}	-1.56
C_2	Electric pump	b_{18}	-7.61
	OWNERSHIP		
O_1	Public	b_{19}	-0.71
O_2	Private individuals and groups	b_{20}	0.01
O_3	Private firm	b_{21}	-0.17

$R^2 = 0.54$

Adjusted $R^2 = 0.54$

Standard error of estimate = 4.74

N = 21,974,000
n = 2,654

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_{11} + b_6SI_2 + b_7X_1 + b_8X_2 + b_9X_3 + b_{10}PS_1 + b_{11}PO_1 + b_{12}PO_2 + b_{13}PO_3 + b_{14}X_4 + b_{15}ST_1 + b_{16}ST_2 + b_{17}C_1 + b_{18}C_2 + b_{19}O_1 + b_{20}O_2 + b_{21}O_3$$

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Table XII-11

Summary Table of Intermediate and Final Models for the Accessibility Index

Variable/Parameter	Final Model	Household Model	Supply Model
Intercept	19.39	17.51	16.88
REGION			
Northeast	-0.47		
North Central	0.61		
South	0.19		
SMSA/NonSMSA			
SMSA	0.15		
SIZE OF PLACE			
Large rural communities	-0.29		
Small rural communities	0.00		
DEVICES	-0.17	-0.61	
OUTDOOR USES	-0.33	-0.61	
EDUCATION	-0.01	-0.03	
PREFERRED SYSTEM	-0.71	-1.53	
Different			
PREFERRED OWNER			
Public	-0.17	-0.40	
Private individuals and groups	-0.74	-2.05	
Private firm	-1.12	-2.41	
TOTAL TREATMENT	-0.35		-0.37
SUPPLY TYPE			
Community supply	-6.68		-7.04
Well	0.15		0.18
CONVEYANCE			
Pressurized public supply	-1.56		-2.04
Electric pump	-7.61		7.94

Table XII-11 continued

Variable/Parameter	Final Model	Household Model	Supply Model
OWNERSHIP			
Public	-0.71		
Private individuals and groups	0.01		
Private firm	-0.17		
R^2	0.54	0.16	0.52
Adjusted R^2	0.54	0.16	0.52
Standard error of estimate	4.74	6.39	4.86

TREATMENT, SUPPLY TYPE, and CONVEYANCE are relatively stable. Stability in the coefficients for these variables implies that they are least affected by the introduction of other attributes into the final model. If the final model were to be decomposed into its components, the three supply attributes would account for the largest proportion of the variance in accessibility, while the other variables would be considered extraneous. In fact, the most parsimonious predictive model would consist only of the accessibility index and the coded vectors for CONVEYANCE. Of the variables in the final model, CONVEYANCE is the only attribute that makes a significant contribution to the explanation of accessibility that is largely independent of other factors.

ANALYSIS OF RELIABILITY

The analysis of reliability is generally structured like the preceding analysis of accessibility. Initially, bivariate models are developed for the same sets of household, supply, and system attributes. Then, if any of the independent variables meet the criterion for retention ($R^2 \geq 0.03$), intermediate models are developed independently for each set of factors. These intermediate models are then combined into a final comprehensive model. Since the reliability of rural water supplies was relatively constant across households, it is possible that no variables will satisfy the criterion and that no multivariate models will be formulated.

Before proceeding with the analysis, it was necessary to transform the index to counteract the effect of its unusual distributional properties. In Chapter VI, where the index was described fully, it was reported that the majority of households were concentrated at the zero point on the index. Consequently, instead of the index being treated as a continuous variable, which violates an assumption of regression, the scores were collapsed into two categories and subjected to discriminant analysis. Households with scores of zero constitute one of the categories, while the other category comprises households with scores of one or more. After this transformation, 25.7 percent of the households are in the latter category, referred to as Group 1; the other 74.3 percent, with scores of zero, comprise Group 0. Note that even though households are assigned values, these values have no quantitative meaning since they simply designate group membership.

Besides affecting the nature of the dependent variable, this modification of the index has some other implications for the analysis. In particular, the statistics for the bivariate models are slightly different from those compiled in the analysis of accessibility. Rather than reporting a measure of association (r) and a coefficient of determination (R^2) for all continuous independent variables, an eta^2 coefficient and means are presented. The eta^2 statistic measures the extent to which the dependent variable is systematically related to both continuous and categorical independent variables in the bivariate models, while the means of the continuous variables simply supplement this information. Accompanying the eta^2 for each categorical variable are the percentage in each category of the independent variable and the proportions of Group 0 and Group 1 households in each category. This information was used to augment the eta^2 since it provided somewhat greater detail about the strength and direction of the relationships.

BIVARIATE RELATIONSHIPS

Each of the bivariate models consisted of the nominal level data on the categories of the reliability index and a single independent variable. Consistent with the organization of this portion of the analysis, these models are developed in sets based on whether the independent variables denote aspects of households, supplies, or systems.

Household attributes

The squared eta coefficients (eta^2) for the models involving continuous household variables and the reliability categories as well as the means for the categories are reported in Table XII-12. Although there are some slight mean differences, nine of the ten eta^2 coefficients are 0.00, and the other is 0.01. Clearly, these results indicate that the variables in the model are not systematically related to reliability and that they do not predict membership in Group 0 or Group 1 on the index. Likewise, since the continuous household variables do not seem to be associated with any systematic variation in reliability, they are not retained for multivariate analysis.

A similar pattern of relationships emerges when the categorical household variables are considered. Table XII-13 contains the summary information and squared eta coefficients for reliability and the various sets of categories included in these variables. The magnitude of these coefficients, which are either 0.00 or 0.01, indicates that information about these categorical household variables does not account for the differences on the reliability index. Consequently, these attributes are also excluded from any multivariate model of reliability.

Supply attributes

Insignificant relationships also characterize the data on reliability and the continuous supply variables. Table XII-14 provides the results of the bivariate models involving each of the four continuous supply variables and the two index categories. Consistent with the small differences in the means, the squared eta coefficients (eta^2), which are all 0.00, signify that none of the variables accounts for even as much as 1 percent of the variation in reliability. This, of course, eliminates them from appearing in any intermediate supply model or final model of reliability.

Table XII-12

Summary Statistics for the Reliability Index According to Continuous Household Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
INCOME	16590.4	18815.5	0.00
EDUCATION	11.7	12.4	0.00
PEOPLE	3.0	3.2	0.00
RESIDENCE	10.6	8.8	0.00
YEARS SERVED	9.7	8.2	0.00
UNITS	1.5	1.7	0.00
LIKED	0.8	0.8	0.00
DISLIKED	0.5	0.5	0.01
DEVICES	8.4	9.3	0.00
OUTDOOR USES	1.7	1.9	0.00
N = 21,974,000 n = 2,654			
QUANTITY	24757.5	25534.4	0.00
N = 5,696,000 n = 603			

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

Table XII-13

Summary Statistics for the Reliability Index According to Categorical Household Variables

Variable and Categories	Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
ETHNICITY				0.00
White	91.4	.74	.26	
Nonwhite	8.6	.73	.27	
TENURE				0.00
Own	79.4	.75	.25	
Rent	20.6	.74	.26	
WORKING				0.00
Employed	70.5	.76	.24	
Not employed	29.5	.74	.26	
PREFERRED SYSTEM				0.01
Different	17.8	.67	.33	
Present	82.2	.76	.24	
PREFERRED SIZE				0.00
Individual	39.6	.75	.25	
Intermediate	10.8	.70	.30	
Community	49.6	.75	.25	
PREFERRED OWNER				0.00
Public	41.0	.76	.24	
Private individuals and groups	51.7	.74	.26	
Private firm	6.3	.69	.31	
Other	1.0	.75	.25	
WILLINGNESS TO PAY				0.01
Less	3.6	.74	.26	
Same	63.3	.77	.23	
More	33.1	.70	.30	

N = 21,974,000
n = 2,654

---

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

Table XII-14

Summary Statistics for the Reliability Index According to Continuous Supply Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
AESTHETIC/ECONOMIC QUALITY	2.1	2.3	0.00
TOTAL TREATMENT	2.3	2.3	0.00
N = 21,974,000 n = 2,654			
WELL DEPTH	37.0	33.0	0.00
WELL AGE	12.8	12.6	0.00
N = 9,751,000 n = 1,300			

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

The categorical supply variables are also unrelated to membership in the two groups of the reliability index. Table XII-15 reports the information on proportions and the squared eta coefficients. Once again, the proportions in the two groups, Group 0 and Group 1, are similar across the sets of categories for any single attribute. The largest differences are for WELL TYPE, where the proportions in Group 0 range from 0.68 to 0.93 and the proportions in Group 1 vary from 0.07 to 0.32. Even these dissimilarities are not substantial enough to generate a coefficient that exceeds the criterion level, since the eta^2 for WELL TYPE is only 0.01. Because the other two coefficients are both 0.00, none of the three attributes in this table was retained for a multivariate model.

System attributes

The bivariate models for the continuous system attributes also show little association with variation in reliability. Table XII-16 contains the squared eta coefficients and the means for Group 0 and Group 1 households for the six continuous system variables selected for this analysis. While the means vary more than they have elsewhere, the squared etas again are either 0.00 or 0.01, and none equals or exceeds the required 0.03 level. Since none of the attributes satisfies this condition, they will not be incorporated into any subsequent models.

Categorical attributes, which are evaluated in Table XII-17, are likewise unrelated to reliability. The squared etas for both OWNERSHIP and SOURCE are 0.00, and the proportions of Group 0 and Group 1 are comparatively uniform for both attributes. These system attributes do not qualify for inclusion into a more comprehensive model of reliability since they do not provide the necessary predictive capability.

In sum, the analysis of reliability concludes at this point because none of the independent variables that were examined at the bivariate level account for membership in the index categories (Group 0 and Group 1). This suggests that the

Table XII-15

Summary Statistics for the Reliability Index According to Categorical Supply Variables

Variable and Categories	Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
SUPPLY TYPE				0.00
Community supply	50.0	0.74	0.26	
Well	44.8	0.76	0.24	
Other	5.2	0.67	0.33	
CONVEYANCE				0.00
Pressurized public supply	47.9	0.74	0.26	
Electric pump	46.9	0.74	0.26	
Other	5.2	0.78	0.22	
N = 21,974,000 n = 2,654				
WELL TYPE				0.01
Dug	12.8	0.68	0.32	
Bored	6.2	0.69	0.31	
Drilled	67.8	0.77	0.23	
Driven	11.8	0.83	0.17	
Jetted	1.4	0.93	0.07	
N = 9,751,000 n = 1,300				

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

Table XII-16

Summary Statistics for the Reliability Index According to Continuous System Variables

Variable	Mean for Group 0*	Mean for Group 1**	eta^2
CONNECTIONS	7226.9	5326.2	0.00
N = 21,974,000 n = 2,654			
OPERATORS	15.6	10.2	0.01
SYSTEM AGE	6.2	6.9	0.00
REVENUE	2178056.0	1573059.0	0.00
REVENUE/CXN	399.7	455.7	0.00
PERFORMANCE	13.7	10.9	0.00
N = 13,209,000 n = 1,461			

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

Table XII-17

Summary Statistics for the Reliability Index According to Categorical System Variables

Variable and Categories	Percent in Category	Proportion in Group 0*	Proportion in Group 1**	eta^2
OWNERSHIP				0.00
Public	40.7	0.76	0.24	
Private individuals and groups	52.3	0.74	0.26	
Private firm	6.6	0.65	0.35	
Other	0.4	0.79	0.21	
N = 21,974,000 n = 2,654				
SOURCE				0.00
Groundwater	58.8	0.72	0.28	
Surface water	18.1	0.74	0.26	
Other	23.1	0.76	0.24	
N = 13,209,000 n = 1,461				

*Group 0 includes households with scores of zero on the reliability index.

**Group 1 includes households with scores of one or more on the reliability index.

variation in reliability is principally random, or that differences are attributable to conditions which the NSA did not consider.

SUMMARY AND CONCLUSIONS

The analysis of availability involved the specification of a series of models that explored relationships between indices of accessibility and reliability and various independent variables. These variables were organized into sets depending upon whether they reflected aspects of households, supplies, or systems. Bivariate models, which consisted of the measure of accessibility or reliability and a single independent variable, were developed to determine the magnitude of their joint associations. Then, if an independent variable met the criterion level of association, it was incorporated into a more complex intermediate or final model. If, however, bivariate relationships were negligible, the modeling of those variables did not proceed beyond the bivariate stage.

The examination of the accessibility index, which was conducted first, detected some significant sources of variation. A total of nine independent variables that emerged from the bivariate segments were combined into the final accessibility model along with three design variables. This final model accounted for 54 percent of the variation in accessibility and suggested, with the results of the other models, that the type of water conveyance was the most important variable. Other attributes of prominence included the number of treatment devices and the type of supply from which the household obtained water. Although this final model was a more accurate predictive device than was anticipated, the intermediate supply model, which accounted for 52 percent of the differences between rural households on the accessibility index, was judged to be more efficient. This efficiency was attributable to the supply model providing approximately the same predictive ability as the final model with substantially fewer independent variables. The enhanced accuracy of the final model was not sufficient to compensate for the number of variables that conveyed redundant information.

Although the analysis of reliability encompassed the same sets of independent variables as appeared in the examination of accessibility, it produced some different results. More specifically, since none of the independent variables showed a simple association with reliability, this component of the analysis was terminated with the bivariate stage. Generally, reliability was uniformly distributed across rural households, and it was not systematically related to any of the variables selected for this particular phase of the NSA investigation.

XIII

Factors Related to Rural Water Conditions: Cost

INTRODUCTION

Cost was addressed in the NSA by examining water bills transmitted to rural households. The water bills contained a variety of information ranging from the quantity consumed to the rate structure used in calculating costs. In the NSA, cost was measured in terms of a monthly charge to the user and in terms of a uniform charge for every thousand gallons consumed. Thus, in order to calculate these two measures of cost, the water bill had to contain the following pieces of information: the cost of the water, the time period covered by the water bill, and the quantity of water consumed at the household. Since the water bills provided the basic data for the cost analysis, it was impossible to evaluate the cost to the rural user residing in households served by individual and intermediate systems, since ordinarily these systems do not bill for their services. Hence, the cost analysis was restricted to households served by community systems.

The analysis presented in Chapter XIII is an attempt to isolate factors which are systematically associated with the cost of rural water supplies. The two measures analyzed here, monthly cost and unit cost, respectively, represent very different dimensions of cost as confirmed by their low statistical association (r = 0.30). The former represents the amount paid for a commodity of value and thus varies according to the consumer's particular preferences. For example, if the consumer chooses to use his water for a low value use such as lawn sprinkling, he implicitly agrees to pay the amount associated with increased usage. Thus the monthly charge will vary directly with the amount of water consumed during that period. In contrast, the unit cost is determined by the utility which provides water to the household, based upon its operating and capital expenditures. Operating expenditures refer to the utility's costs for support services, acquisition, treatment, and distribution (including storage), though any one particular water supply system may not have expenditures in all four operating cost categories. Capital expenditures generally include depreciation and interest for the plant during its operating life. Depreciation is typically based on the original cost of the facility, and not on the costs required to reproduce the facility. Interest costs are the amounts utilities must pay for their government loans, bonds, or other financing mechanisms.

Information about household water costs used in this analysis was presented in detail in Chapter V of this report. As explained there, a household's water bills provided the basic NSA data on the costs associated with residential demand. Since water bills covered periods ranging from a single month to an entire year, monthly costs in the latter case generally represented an average cost per month for the year, while the cost associated with a single month's consumption was likely an overestimation, since most NSA interviews were conducted during summer months when water consumption (and therefore water costs) was generally at a peak.

As explained in Chapter V, and also Chapter XI, although 9.2 million rural households received regular water bills, bills were not always available at the time of the NSA interview. Sometimes, too, a bill was available but lacked major components of information important to the NSA analysis. Thus, the data base for each analysis in this chapter represented roughly 5.4 million rural households across the US. The observations (rural households) used in the analysis of monthly cost and unit cost were not necessarily the same, though there was a high degree of overlap. The data base used in the analysis of monthly cost contained a total of 571 observations, while the data base for unit cost comprised 567 sampled households.

The two dependent variables, monthly cost and unit cost, are analyzed separately in the sections to follow. The format established for the earlier analytical chapters is again adopted here. First, in each major section, the independent variables are selected and examined, and then bivariate relationships are explored. Intermediate models are developed next, culminating in a final model from which conclusions are drawn.

ANALYSIS OF MONTHLY COST

HOUSEHOLD ATTRIBUTES

Since monthly household water costs are determined, in part, by the amount of water consumed during a given month, it was expected that many of the sociodemographic characteristics found to be associated with recorded quantity, in addition to that variable itself, would also relate to monthly household costs. For example, two characteristics which partially accounted for variation in residential water demand—the size of the household and the number of water-using fixtures and devices—were thought to influence monthly costs through their association with recorded quantity. Though household income was not retained for inclusion in the demand model, that variable may determine in large measure monthly allocations for water.

Certain other relationships were also explored—various characteristics associated with the household head, rural residents' perceptions of water supply conditions, and outdoor water usage. Household attributes studied in the NSA analysis of cost are listed in Table XIII-1. Though this same set of variables was explored for both dimensions of cost, the expectation was that the analysis would identify a very different set of factors for each of the cost measures.

Because of the reduced case base in the cost analysis, it was necessary to treat one of the independent variables somewhat differently than was described in Chapter IX (see Table IX-1). Specifically, categories for employment status (WORKING) were reduced to two, employed and not employed. The employed group consisted of household heads who were working either full time or part time. The unemployed group consisted of household heads who were keeping house, going to school, retired, or unemployed.

SUPPLY ATTRIBUTES

The cost of a commodity has traditionally been equated with the quality of the product. In terms of water supply, then, the superiority of the product encompasses a number of quality-related features, including reliability of the supply, accessibility, and physical, microbiological, and chemical properties of the water supply. Also, the extent of water treatment may contribute to the cost of the final product.

The supply attributes examined in the cost analysis are listed in Table XIII-1. They are described in Chapter IX (see Table IX-2); no modifications specific to the cost analysis were necessary.

SYSTEM ATTRIBUTES

Certain system attributes also may influence the monthly cost of water. Previous research on the costs associated with water production indicates that a wide variety of factors ranging from the size of the system to the source from which water is withdrawn are incorporated in the charges attached to revenue-producing water. System ownership has a major influence on water charges to the

Table XIII-1

Independent Variables Used in the Analysis of Unit Cost and Monthly Cost

Continuous Variables	Categorical Variables
Household Attributes	
AGE	ETHNICITY (white, nonwhite)
EDUCATION	MOBILE HOME (mobile home, not mobile home)
INCOME	PERCEIVED COST (inexpensive, reasonable, expensive, very expensive)
PEOPLE	TENURE (own, rent, other)
RESIDENCE	WILLINGNESS TO PAY (less, same, more)
UNITS	WORKING (employed, not employed)
YEARS SERVED	PREFERRED SYSTEM (present system, different system)
DEVICES	
DISLIKED	
LIKED	
PERCEIVED QUALITY	
SATISFACTION	
QUANTITY	
OUTDOOR USES	
Supply Attributes	
ACCESSIBILITY	None
AESTHETIC/ECONOMIC QUALITY	
RELIABILITY	
TOTAL TREATMENT	
System Attributes	
CONNECTIONS	OWNERSHIP (public, private individuals and groups, private firm)
DEBT RETIREMENT	SOURCE (groundwater, surface water, purchased, mixed)
LINES	
O & M	
OPERATORS	
REVENUE	
REVENUE/CXN	
SYSTEM AGE	
PERFORMANCE	

user. However, the rural household is one step removed from the system which supplies its water, and therefore monthly expenditures for water are likely influenced by not only water system attributes, but by characteristics specific to the household and attributes of the water supply.

Water system attributes examined in the cost analysis are listed in Table XIII-1. They are described in Chapter IX (see Table IX-3). Though modifications were not necessary, forms of ownership in the "other" category were not encountered because of the reduced case base.

BIVARIATE RELATIONSHIPS

Distribution of monthly cost

Water costs calculated on a monthly basis varied considerably among rural households. These costs ranged from less than one dollar a month through $198. The distribution was very highly skewed as indicated by the average cost of about nine dollars per month. The standard deviation of the distribution was $10.14, reflecting the wide variation in monthly cost.

Initial examination of a few bivariate relationships involving monthly cost showed that the relationships were essentially curvilinear in form, rather than linear. Therefore, to accomplish the analysis using the standard linear regression model, a logarithmic transformation was applied to monthly cost. Besides linearizing its relationship to a number of independent variables, the transformation produced a more normal univariate distribution. The transformed variable ranged from -0.18 through 5.29. The mean was 1.94, while the standard deviation of the distribution was 0.69. The negative values of the transformed variable correspond to a cost less than one dollar; a value of zero is associated with a cost of exactly one dollar; and the positive range of values on the dependent variable correspond to costs which exceed one dollar.

Household attributes

Data pertaining to household attributes and monthly water costs are presented in Tables XIII-2 and XIII-3. Table XIII-2 presents the correlations found between cost and the continuous variables which were examined. Table XIII-3 presents summary statistics for categorical variables. Six attributes—the quantity consumed (QUANTITY), the number of household occupants (PEOPLE), the number of water-using fixtures and devices (DEVICES), the number of years served by the water system (YEARS SERVED), the education of the household head (EDUCATION), and the respondent's perceptions concerning the cost of the household water supply (PERCEIVED COST)—met the criterion used to determine analytical significance (an R^2 value of at least 0.03).

The quantity consumed at the household on a monthly basis was the single most important factor in estimating household water costs. Increases in monthly expenditures for water were associated with increased consumption, as reflected in the correlation coefficient ($r = 0.53$). The variable QUANTITY accounted for fully 28 percent of the variance in the dependent measure.

As one might expect, there was a positive relationship between household size and water costs—as the number of occupants increased, so did monthly water costs. As shown in Table XIII-2, the strength of the relationship was moderate ($r = 0.33$).

Similarly, monthly water costs were positively associated with the number of water-using fixtures and devices at the household. The correlation was somewhat lower ($r = 0.21$), though in the expected direction.

The correlation between monthly household water costs and the length of time served by the present water system was negative, and the strength of the association was slight as indicated by the correlation coefficient ($r = -0.20$). Though the finding is somewhat difficult to interpret, the relationship indicates a decline in household water expenditures over the years. Perhaps one explanation is

Table XIII-2

Correlations Between Monthly Cost and Continuous Household Variables

Variable	r	R^2
AGE	-0.10	0.01
EDUCATION	0.18	0.03*
INCOME	0.16	0.03
PEOPLE	0.33	0.11
RESIDENCE	-0.17	0.03
YEARS SERVED	-0.20	0.04
UNITS	-0.14	0.02
PERCEIVED QUALITY	0.02	0.00
SATISFACTION	-0.02	0.00
LIKED	0.08	0.01
DISLIKED	0.04	0.00
DEVICES	0.21	0.04
OUTDOOR USES	0.11	0.01
QUANTITY	0.53	0.28

N = 5,395,000
n = 571

*Underscoring denotes that the variable was retained for further analysis.

Table XIII-3

Summary Statistics for Monthly Cost According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
ETHNICITY				0.01
White	88.8	1.92	0.66	
Nonwhite	11.2	2.12	0.68	
MOBILE HOME				0.00
Not mobile home	90.8	1.94	0.68	
Mobile home	9.2	1.92	0.55	
TENURE				0.01
Own	89.0	1.96	0.65	
Rent	10.3	1.76	0.76	
Other	0.8	2.11	0.61	
WORKING				0.02
Employed	69.1	2.00	0.68	
Not employed	30.9	1.80	0.61	
PERCEIVED COST				0.14*
Inexpensive	12.0	1.57	0.64	
Reasonable	65.3	1.89	0.61	
Expensive	15.9	2.10	0.54	
Very expensive	6.8	2.70	0.86	
PREFERRED SYSTEM				0.01
Present system	84.0	1.91	0.67	
Different system	16.0	2.09	0.65	
WILLINGNESS TO PAY				0.02
Less	2.1	2.20	0.59	
Same	64.7	1.88	0.65	
More	33.2	2.33	0.72	

N = 5,395,000
n = 571

*Underscoring denotes that the variable was retained for further analysis.

that costs are lower among older residents in accordance with their reduced consumption.

Education of the household head just met the minimum criterion established for the study ($R^2 = 0.03$). The correlation coefficient indicated that monthly water costs were higher among households with more educated heads. Indeed, this finding corroborated previous research results by Spaulding who suggested that demand for water (and therefore associated costs) may be related to certain life styles which vary according to education and income.[1] As shown in Table XIII-2, the relationship between costs and income was in the direction that Spaulding indicated, but was not retained for further analysis since the R^2 value (table values are rounded) fell slightly below the criterion for selecting variables to include in the analytical model.

Finally, the respondent's notions about the cost of the water supply closely matched the actual amount paid for water during a given month. As shown in Table XIII-3, this household attribute accounted for fully 14 percent of the variability in monthly water costs. The mean values for the transformed variable increased as perceived water costs climbed. (Since the means reported for categorical independent variables represent the score on the transformed dependent variable for a given category, interpretation is limited to simply the direction of the relationship, as indicated by increasing or decreasing mean values for the categories).

Other household attributes did not show any important systematic relationships with the dependent variable under investigation, as indicated by the generally low R^2 values. One exception, however, was the variable RESIDENCE which shows a R^2 value of 0.03, but this figure was rounded and did not actually meet the criterion established for the study.

Supply attributes

Supply attributes were relatively unimportant in their associations with monthly household water costs. Only supply reliability and the aesthetic/economic water quality index accounted for even 1 percent of the variability in the dependent measure (see Table XIII-4).

System attributes

Like supply attributes, system attributes showed almost no systematic relationships with monthly household water costs (see Tables XIII-5 and XIII-6). Economic attributes showed very low correlations with monthly cost as did physical and operational characteristics of the supply system. One variable, however, the source from which the water was withdrawn (SOURCE), fell just slightly below the R^2 criterion, and so was excluded from further consideration (table values are rounded).

System ownership (OWNERSHIP) accounted for roughly 5 percent of the variability in monthly water costs and therefore was retained for final model specification (see Table XIII-6). This finding was consistent with the results presented in Chapter XI; that is, this same variable accounted for approximately the same proportion of the variability in household water consumption. This was not surprising given the strong association between household water consumption and monthly water costs.

The pattern conveyed by the average scores for each ownership category was in agreement with pricing policies utilized by different ownership groups. Since private systems need to realize some profit, they may charge more for water than public systems who often depend upon subsidies from general municipal funds.[2] As shown in Table XIII-6, the average transformed score was considerably higher among households served by private systems than among those served by public systems (2.33, as compared to 1.88).

Table XIII-4

Correlations Between Monthly Cost and Continuous Supply Variables

Variable	r	R^2
ACCESSIBILITY	0.03	0.00
RELIABILITY	0.07	0.01
AESTHETIC/ECONOMIC QUALITY	0.09	0.01
TOTAL TREATMENT	0.05	0.00

N = 5,395,000
n = 571

Table XIII-5

Correlations Between Monthly Cost and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.06	0.00
LINES	-0.02	0.00
SYSTEM AGE	0.09	0.01
OPERATORS	0.00	0.00
REVENUE	0.01	0.00
REVENUE/CXN	0.05	0.00
O & M	-0.02	0.00
DEBT RETIREMENT	0.07	0.01
PERFORMANCE	0.08	0.01

N = 5,395,000
n = 571

Table XIII-6

Summary Statistics for Monthly Cost According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SOURCE				0.03
Groundwater	49.9	1.90	0.70	
Surface water	22.3	1.82	0.58	
Purchased	10.8	2.13	0.52	
Mixed	16.9	2.10	0.71	
OWNERSHIP				0.05*
Public	85.4	1.88	0.65	
Private firm	10.0	2.33	0.72	
Private groups and individuals	4.7	2.20	0.59	

N = 5,395,000
n = 571

*Underscoring denotes that the variable was retained for further analysis.

INTERMEDIATE MODEL SPECIFICATION

The second stage of the analysis was to develop a separate regression model for each of the three groups of attributes in order to portray the overall impact of each set on the dependent variable, monthly water costs. Each model included only those attributes which met the criterion of accounting for at least 3 percent of the variability in cost in the bivariate analysis. Among household attributes, six had met the criterion: the quantity of water consumed during a month (QUANTITY), the size of the household (PEOPLE), the number of water-using fixtures and devices at the household (DEVICES), the length of time served by the present water system (YEARS SERVED), the educational attainment of the household head (EDUCATION), and perceived cost (PERCEIVED COST). No supply attributes met the analysis criterion. One system attribute met the criterion—the type of ownership of the system (OWNERSHIP).

Intermediate household model

The first step in the multivariate analysis was to assess the multicollinearity of the items to be considered. Zero-order correlations among the items were generally low (below 0.25), except in the case of water-using devices and education. Though the correlation between these items was strong as indicated by the correlation coefficient of 0.51, collinearity problems are generally associated with correlations of a larger magnitude, 0.70 or higher. Consequently, all variables were retained for the intermediate model.

Household attributes were more important in their association with monthly water costs than either supply attributes or system attributes. This was evidenced by the results of the linear regression model, as shown in Table XIII-7. Specifically, the model accounts for 42 percent of the variability in water costs.

Apart from QUANTITY, which was the single most important factor in determining monthly expenditures for water, household size (PEOPLE) retained its

Table XIII-7

Intermediate Regression Estimates for Monthly Cost and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	1.313
X_1	QUANTITY (in thousands)	b_1	0.013*
	PERCEIVED COST		
PC_1	Inexpensive	b_2	-0.458
PC_2	Reasonable	b_3	-0.091
PC_3	Expensive	b_4	0.102
X_2	PEOPLE	b_5	0.059
X_3	DEVICES	b_6	0.002
X_4	YEARS SERVED	b_7	-0.004
X_5	EDUCATION	b_8	0.019

$R^2 = 0.42$

Adjusted $R^2 = 0.42$

Standard error of estimate = 0.509

N = 5,395,000
n = 571

$$Y' = a + b_1X_1 + b_2PC_1 + b_3PC_2 + b_4PC_3 + b_5X_2 + b_6X_3 + b_7X_4 + b_8X_5$$

*QUANTITY is reported in thousands of liters. Therefore, to obtain an unstandardized estimate for the unadjusted variable, divide the estimate by 1,000.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

superior importance in the context of the multiple variable approach as compared to other continuous independent variables included in the model. The estimated regression coefficient for PEOPLE was significantly larger than the estimates for DEVICES, YEARS SERVED, and EDUCATION (see Table XIII-7). The number of water-using fixtures and devices installed at the household was least important, as judged by the magnitude of its coefficient. (Again, the regression coefficients are reported for the transformed dependent variable and therefore must be interpreted as a change in the natural log of dollars spent for water per month).

Intermediate supply model

Since the bivariate analysis yielded no systematic patterns of association between monthly water costs and the supply attributes, intermediate model construction was precluded.

Intermediate system model

Since only one system attribute was retained for inclusion in the inter- diate model (OWNERSHIP), a regression estimate would merely duplicate the rmation from the bivariate analysis. Consequently, the information is not peated here.

FINAL MODEL SPECIFICATION

The final model of factors related to household water costs incorporated system ownership and the household attributes which appeared in the intermediate model. Also, for reasons discussed in Chapter IX, the sampling design variables (region, SMSA/nonSMSA, and size of place) were also included in the final model.

As illustrated in Table XIII-8, the final model yielded only slight improvement over the simpler intermediate household model. The amount of variance accounted for by the final model was 47 percent, a gain of 5 percent over the

Table XIII-8

Final Regression Estimates for Monthly Cost

Variable Designation		Parameter	Estimate
	Intercept	a	1.343
	REGION		
R_1	Northeast	b_1	-0.169
R_2	North Central	b_2	-0.054
R_3	South	b_3	0.018
	SMSA/NonSMSA		
S_1	SMSA	b_4	-0.066
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.034
SI_2	Small rural communities	b_6	-0.006
X_1	QUANTITY (in thousands)	b_7	0.012*
	PERCEIVED COST		
PC_1	Inexpensive	b_8	-0.467
PC_2	Reasonable	b_9	-0.096
PC_3	Expensive	b_{10}	0.122
X_2	PEOPLE	b_{11}	0.055
	OWNERSHIP		
O_1	Private groups and individuals	b_{12}	0.015
O_2	Public	b_{13}	-0.127
X_3	DEVICES	b_{14}	0.002
X_4	YEARS SERVED	b_{15}	-0.003
X_5	EDUCATION	b_{16}	0.022

$R^2 = 0.47$

Adjusted $R^2 = 0.47$

Standard error of estimate = 0.488

Table XIII-8 continued

Variable Designation	Parameter	Estimate
N = 5,395,000 n = 571		

$$\begin{aligned} Y' &= a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + \\ &\quad b_7X_1 + b_8PC_1 + b_9PC_2 + b_{10}PC_3 + b_{11}X_2 + b_{12}O_1 + \\ &\quad b_{13}O_2 + b_{14}X_3 + b_{15}X_4 + b_{16}X_5 \end{aligned}$$

*QUANTITY is reported in thousands of liters. Therefore, to obtain an unstandardized estimate for the unadjusted variable, divide the estimate by 1,000.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

intermediate household model. The standard error of estimate was reduced by 4 percent, from 0.509 to 0.488.

When direct comparisons were made of the intermediate and final models, the magnitudes of the estimated regression coefficients remained relatively stable across models. This finding supports the earlier statement that the inclusion of form of system ownership and the design variables had little impact on the predictive power of the model. However, some regional difference in monthly water costs was evident in the regression, as indicated by the differences in the magnitudes of their respective regression coefficients (see Table XIII-8). More specifically, monthly allocations for domestic water were lowest in the Northeast and highest in the West.

ANALYSIS OF UNIT COST

HOUSEHOLD ATTRIBUTES

The previous analysis identified one conceptual set of variables—those associated with the household—to be particularly important in determining monthly expenditures for water. Here, however, in the analysis of unit cost, the expectation was that household attributes would have very little systematic association with the dependent variable under investigation. This belief was based on knowledge of various pricing strategies utilized by the water supply industry. One exception, however, as indicated in Chapter XI, was the fairly strong association between the unit cost (or cost per thousand gallons of water) and the quantity consumed, both attributes of the household. This finding relates to well-established pricing policies which offer a reduced charge for additional units of water consumed (decreasing block and step rates).

The set of household attributes studied in the analysis of monthly cost are explored here as well (see Table XIII-1). As before, categories for the variable WORKING were reduced to two, employed and not employed. The employed group consisted of household heads who were working either full time or part time, while the unemployed group consisted of heads who were keeping house, going to school, retired, or unemployed.

One final point is worthy of mention, though it did not involve a variable modification. The units of measurement differed on the quantity consumed at the household and the unit cost of water, one bivariate relationship explored here. It was important to report quantity in liters, since that is how it appeared throughout the report (see Chapters V, XI, and XII). On the other hand, unit cost was reported in dollars per thousand gallons (see Chapters V and XI), since that is one convention used by the water supply industry.

SUPPLY ATTRIBUTES

Various characteristics of the water supply may have a bearing on the unit cost of water. A tendency towards improved water quality conditions, for instance, may necessitate an increase in the unit cost of water. Likewise, a more accessible or reliable water supply may require special equipment or increased operating personnel to provide the improved service.

The supply attributes investigated in the analysis of monthly household costs are explored here also. They are listed in Table XIII-1.

SYSTEM ATTRIBUTES

System attributes were expected to influence unit costs more directly than either household attributes or supply attributes. Efficiencies achieved through increased size or "economies of scale" lend support to this argument. Expenses associated with water production which, in part, determine the cost to the user, may vary according to the number of system operators or the types and extent of water treatment at the system. Moreover, the economics of the water system may also determine the unit cost of the water provided.

These, and other relationships, are explored in the following analysis. The system attributes, examined in the preceding analysis and here too, are listed in Table XIII-1.

BIVARIATE RELATIONSHIPS

Distribution of unit cost

Unit water costs varied greatly among rural households, ranging from $0.13 through $78.17 per thousand gallons. The distribution was highly skewed as indicated by the average cost of $2.50, with a standard deviation of $5.25.

Preliminary examination of the univariate distribution for unit cost and certain bivariate relationships indicated the necessity of transforming the variable so that the assumptions of linear regression were met. Similar to the analysis of monthly cost, a logarithmic transformation was applied to UNIT COST. The transformation produced a distribution which ranged from -2.03 through 4.36. The mean and standard deviation of this distribution was 0.43 and 0.84, respectively.

Household attributes

Data pertaining to household attributes and unit water costs are presented in Tables XIII-9 and XIII-10. Table XIII-9 presents the correlations and R^2 values found between unit cost and the continuous variables which were examined. Table XIII-10 presents summary statistics including eta^2 for categorical variables.

Generally speaking, household attributes showed little association with unit cost, except for recorded quantity and perceived cost. The cost/quantity relationship was previously described in Chapter XI. In brief, recorded quantity accounted for fully 22 percent of the variance in the dependent measure. The association between unit cost and perceived cost was considerably lower (R^2 = [illegible]), though it met the criterion established for the study. The association was positive, indicating simply that higher actual costs were associated with higher perceived

Table XIII-9

Correlations Between Unit Cost and Continuous Household Variables

Variable	r	R^2
AGE	0.12	0.01
EDUCATION	-0.02	0.00
INCOME	-0.09	0.01
PEOPLE	-0.13	0.02
RESIDENCE	0.06	0.00
YEARS SERVED	-0.00	0.00
UNITS	-0.09	0.01
PERCEIVED QUALITY	0.09	0.01
SATISFACTION	-0.14	0.02
LIKED	-0.01	0.00
DISLIKED	0.10	0.01
QUANTITY	-0.47	0.22*
OUTDOOR USES	-0.03	0.00

N = 3,374,000
n = 367

---

*Underscoring denotes that the variable was retained for further analysis.

Table XIII-10

Summary Statistics for Unit Cost According to Categorical Household Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
ETHNICITY				0.00
White	88.7	0.41	0.84	
Nonwhite	11.3	0.53	0.90	
MOBILE HOME				0.00
Not mobile home	90.8	0.50	0.67	
Mobile home	9.2	0.42	0.86	
TENURE				0.01
Own	88.9	0.45	0.82	
Rent	10.3	0.21	0.96	
Other	0.8	0.97	0.81	
WORKING				0.00
Employed	69.1	0.39	0.86	
Not employed	30.9	0.51	0.80	
PERCEIVED COST				0.05*
Inexpensive	12.0	0.01	0.66	
Reasonable	65.3	0.42	0.88	
Expensive	15.9	0.63	0.81	
Very expensive	6.8	0.73	0.38	
PREFERRED SYSTEM				0.02
Present system	84.0	0.37	0.80	
Different system	16.0	0.72	0.99	
WILLINGNESS TO PAY				0.02
Less	2.1	1.10	0.91	
Same	64.6	0.39	0.83	
More	33.2	0.46	0.83	

N = 3,376,000
n = 567

---

*Underlining denotes that the variable was retained for further analysis.

costs. The means for each of the perceived cost categories are shown in Table XIII-10. Since they are reported for the transformed dependent variable, no interpretation other than the direction of the relationship should be made.

Supply attributes

As shown in Table XIII-11, of the four supply attributes examined in the analysis of unit cost only one—the quality of the water supply with respect to constituents having aesthetic or economic effects (AESTHETIC/ECONOMIC QUALITY)—showed a systematic association with the dependent measure. The correlation was slight ($r = 0.18$) and showed that supplies which were of higher quality (as indicated by low scores on the aesthetic/economic index) tended to show lower unit costs.

System attributes

Data pertaining to system attributes and unit water costs are presented in Tables XIII-12 and XIII-13. Table XIII-12 presents the correlations between cost and the continuous variables which were examined. Table XIII-13 presents summary statistics including eta^2 for categorical variables.

An examination of many bivariate relationships involving system attributes revealed that only two variables in the set were worthy of further analysis. In concert with established theory, certain economies of scale were achieved, as indicated by the correlation between the number of service connections and the unit cost of water ($r = -0.19$). Simply stated, the cost of 1,000 gallons of water over the entire quantity consumed was generally lower among households served by large community water systems. The source from which the water is withdrawn also met the .05 criterion used to establish statistical significance. As shown in Table XIII-13, this variable accounted for roughly 3 percent of the variability in unit water costs. The mean scores on the natural log of unit cost are reported for

Table XIII-11

Correlations Between Unit Cost and Continuous Supply Variables

Variable	r	R^2
ACCESSIBILITY	-0.04	0.00
RELIABILITY	0.03	0.00
AESTHETIC/ECONOMIC QUALITY	0.18	0.03*
TOTAL TREATMENT	0.08	0.01

N = 5,374,000
n = 567

---

*Underscoring denotes that the variable was retained for further analysis.

Table XIII-12

Correlations Between Unit Cost and Continuous System Variables

Variable	r	R^2
CONNECTIONS	-0.19	0.03*
LINES	-0.11	0.01
SYSTEM AGE	-0.03	0.00
OPERATORS	-0.12	0.01
REVENUE	-0.15	0.02
REVENUE/CXN	-0.02	0.00
O & M	-0.04	0.00
DEBT RETIREMENT	0.01	0.00
PERFORMANCE	-0.00	0.00

N = 5,374,000
n = 567

*Underscoring denotes that the variable was retained for further analysis.

Table XIII-13

Summary Statistics for Unit Cost According to Categorical System Variables

Variable and Categories	Percent in Category	Mean	Standard Deviation	eta^2
SOURCE				0.05*
Groundwater	49.9	0.25	0.80	
Surface water	22.3	0.48	0.74	
Purchased	10.8	0.69	0.73	
Mixed	16.9	0.70	1.02	
OWNERSHIP				0.00
Public	85.3	0.43	0.82	
Private firms	10.0	0.40	0.97	
Private groups and individuals	4.7	0.55	0.95	

N = 5,374,000
n = 567

---

*Underscoring denotes that the variable was retained for further analysis.

each of the source categories. These scores indicate that the cost of 1,000 gallons of water is generally lower among households served by systems which use a groundwater source, as opposed to those which purchase their water or use multiple sources.

INTERMEDIATE MODEL SPECIFICATION

The second stage of the analysis was to develop a separate regression model for each of the three groups of attributes. Each model included only those attributes which met the criterion of accounting for at least 3 percent of the variability in cost ($R^2 = 0.03$) in the bivariate analysis. Among household attributes, two had met the criterion: the quantity consumed at the household (QUANTITY) and the respondents' perceptions about the cost of the water supply (PERCEIVED COST). With regard to supply attributes, only one variable, a water quality index based on laboratory measurements of constituents having aesthetic or economic effects (AESTHETIC/ECONOMIC QUALITY) was retained for further analysis. Finally, two system attributes met the criterion—the size of the system as determined by the number of service connections (CONNECTIONS) and the system's water source (SOURCE).

Intermediate household model

The results of the regression analysis on household attributes, as well as the formal expression for the equation, are shown in Table XIII-14. The model accounts for 29 percent of the variance in unit cost, as indicated by the R^2 value.

The intent of this phase of the analysis was not to interpret the unstandardized regression estimates but to, first, assess the relative importance of each of the three groups of attributes. Though no comparisons can be made at this point, the household model performed fairly well in terms of its overall effectiveness in accounting for a substantial proportion of the variance in unit water costs.

Table XIII-14

Intermediate Regression Estimates for Unit Cost and Household Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	0.959
X_1	QUANTITY (in thousands)	b_1	-0.018*
	PERCEIVED COST		
PC_1	Inexpensive	b_2	-0.500
PC_2	Reasonable	b_3	-0.098
PC_3	Expensive	b_4	0.088

$R^2 = 0.29$

Adjusted $R^2 = 0.29$

Standard error of estimate = 0.713

N = 5,374,000
n = 567

$$Y' = a + b_1X_1 + b_2PC_1 + b_3PC_2 + b_4PC_3$$

*QUANTITY is reported in thousands of liters. Therefore, to obtain an unstandardized estimate for the unadjusted variable, divide the estimate by 1,000.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

Secondly, the direction of each of the relationships (indicated by the sign of the regression coefficient) was important in the context of all variables included in the intermediate model. As shown in Table XIII-14, the association between unit cost and quantity was similar to the simple bivariate relationship. That is, an increase in household consumption was associated with a decrease in unit cost. Furthermore, average scores on unit cost for households which reported inexpensive and reasonable water costs were less than the unweighted average score on unit cost for all perceived cost categories. Again, this interpretation is consistent with the bivariate findings.

Intermediate supply model

Since only one supply attribute was retained for inclusion in the intermediate model (AESTHETIC/ECONOMIC QUALITY), a regression estimate would merely duplicate the information from the bivariate analysis. Therefore, the information is not repeated here.

Intermediate system model

The intermediate model of factors related to the water supply system is shown in Table XIII-15. As compared to the intermediate household model, system attributes, considered as a group, were generally less important in their association with unit water costs. The model accounts for 13 percent of the variance in unit cost, considerably less than that observed for the set of household factors.

As was found at the bivariate stage of analysis, the size of the water system was inversely related to the unit cost of water. A comparison of source water categories showed that only unit water costs among households served by systems which used a groundwater source were less than the unweighted average unit cost among households served by systems which used all types of source water.

Table XIII-15

Intermediate Regression Estimates for Unit Cost and System Attributes

Variable Designation		Parameter	Estimate
	Intercept	a	0.683
	SOURCE		
SO_1	Groundwater	b_1	-0.384
SO_2	Surface water	b_2	0.020
SO_3	Purchased	b_3	0.053
X_1	CONNECTIONS (in thousands)	b_4	-0.009*

$R^2 = 0.13$

Adjusted $R^2 = 0.13$

Standard error of estimate = 0.789

N = 5,374,00
n = 567

$$Y' = a + b_1SO_1 + b_2SO_2 + b_3SO_3 + b_4X_1$$

*CONNECTIONS is reported in thousands. Therefore, to obtain an unstandardized estimate for the unadjusted variable, divide the estimate by 1,000.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

FINAL MODEL SPECIFICATION

The final model of factors related to unit water costs incorporated the aesthetic/economic water quality index and the household and system attributes which appeared in the intermediate models. Also, the design variables were included in the final model because of their importance in the sample selection process.

The combined effect of each of the three sets of factors included in the final model showed a marked improvement over separate considerations of household and system attributes. More specifically, as shown in Table XIII-16, the model accounted for nearly 40 percent of the variance in cost per thousand gallons of water consumed at rural households across the US.

There was practically no difference in unit water costs between SMSA and nonSMSA areas, and among the three size of place groupings as indicated by their respective regression estimates which clustered about zero. In contrast, however, when statistically controlling for all variables included in the model, regional differences did prevail. More specifically, unit water costs were lowest in the Northeast, and highest in the West.

SUMMARY AND CONCLUSIONS

The purpose of the analyses presented in this chapter was to identify sets of factors which showed patterns of association with monthly household water costs and unit water costs. Among the many factors explored were household attributes, supply attributes, system attributes, and variables used to generate the NSA sample.

Cost was addressed in the NSA by examining water bills transmitted to rural households. This meant, however, that there was no reliable way to collect cost information for households which had their own wells, cisterns, or surface supplies. Moreover, after the data collection activities were complete and the data

Table XIII-16
Final Regression Estimates for Unit Cost

Variable Designation		Parameter	Estimate
	Intercept	a	1.052
	REGION		
R_1	Northeast	b_1	-0.241
R_2	North Central	b_2	-0.075
R_3	South	b_3	0.027
	SMSA/NonSMSA		
S_1	SMSA	b_4	0.082
	SIZE OF PLACE		
SI_1	Large rural communities	b_5	-0.050
SI_2	Small rural communities	b_6	-0.014
X_1	QUANTITY (in thousands)	b_7	-0.017*
	PERCEIVED COST		
PC_1	Inexpensive	b_8	-0.441
PC_2	Reasonable	b_9	-0.107
PC_3	Expensive	b_{10}	0.081
	SOURCE		
SO_1	Groundwater	b_{11}	-0.183
SO_2	Surface water	b_{12}	0.021
SO_3	Purchased	b_{13}	-0.067
X_2	AESTHETIC/ECONOMIC QUALITY	b_{14}	0.023
X_3	CONNECTIONS (in thousands)	b_{15}	-0.008*

$R^2 = 0.39$

Adjusted $R^2 = 0.39$

Standard error of estimate = 0.657

Table XIII-16 continued

Variable Designation	Parameter	Estimate
N = 5,374,000 n = 567		

$$Y' = a + b_1R_1 + b_2R_2 + b_3R_3 + b_4S_1 + b_5SI_1 + b_6SI_2 + b_7X_1 + b_8PC_1 + b_9PC_2 + b_{10}PC_3 + b_{11}SO_1 + b_{12}SO_2 + b_{13}SO_3 + b_{14}X_2 + b_{15}X_3$$

*QUANTITY and CONNECTIONS are reported in thousands. Therefore, to obtain unstandardized estimates for the unadjusted variables, divide the estimates by 1,000.

The unstandardized regression coefficients for effect-coded categorical variables represent the adjusted difference between the identified category and the unweighted average score on the dependent variable for all categories contained in the variable. All coefficients are adjusted for the other independent variables in the equation.

stored on magnetic tape, it was found that similar information for households served by intermediate systems was also lacking. Few of these systems billed their customers and those that did generally did not provide written documentation of the transaction. Hence, the generalizability of the associations presented herein are somewhat restricted and pertain specifically to households served by community water systems.

The analysis of monthly household water costs identified a number of important factors, but the dominant factor was obviously the amount of water consumed during a single month. Variables of lesser importance which showed associations with monthly costs included the number of household occupants, the number of water-using fixtures and devices installed at the household, the number of years served by the water system, the education of the household head, and the respondent's perceptions concerning the cost of the household water supply. All of the variables were positively associated with the dependent measure, excluding the length of time served by the water system. A positive association, between household size and water costs, for instance, indicated that as the number of people residing in the household increased, so did their monthly water expenditures. Conversely, monthly water costs were lower among elderly rural dwellers. In summary, variables relating to the household were considerably more important in determining monthly water costs than either variables identified with the supply or those associated with the water system.

A similar set of factors emerged from the analysis of unit water costs, but a number of important differences also surfaced. As to the similarities of the two analyses, the amount of water consumed and the respondent's perceptions concerning the cost of the supply were also identified as important correlates of unit water costs, although the strength of the relationships was somewhat mitigated by comparison. While an increase in water consumption was associated with an increase in monthly water expenditures, the reverse was true for recorded quantity and unit costs. This finding supports the standard postulate of economic theory which asserts that higher prices will result in reduced consumption.

Also in accordance with economic theory, certain economies of scale were achieved as evidenced by the relationship between the number of service connections and the unit cost of water. This meant that the cost of every 1,000 gallons of water consumed was generally lower among households served by large community systems. Another finding of importance was the association between the source of water and unit costs. Households which purchased water from systems which used a groundwater source, paid considerably less per unit of water consumed than those purchasing from systems which used a surface source. In summary, household attributes and water system attributes were equally important in establishing the unit cost of water.

REFERENCES

[1] Spaulding, Irving A. "Social Class and Household Water Consumption." Social Behavior, Natural Resources, and the Environment. New York: Harper & Row, 1972.

[2] Seidel, Harris F. and Robert E. Baumann. "A Statistical Analysis of Water Works Data for 1955." Journal of the American Water Works Association XLIX (December 1957): 1548.

Bibliography

Allen, Martin J., and Geldreich, Edwin E. "Bacteriological Criteria for Ground-Water Quality." Ground Water 13 (January-February 1975): 45-52.

American Association for Vocational Instructional Materials. Planning for an Individual Water System. Athens, Georgia: American Association for Vocational Instructional Materials, May 1973.

American Heart Association. "Impact of the Environment on Cardiovascular Disease: Report of the American Heart Association Task Force on Environment and Cardiovascular System." 1980.

American Public Health Association, American Water Works Association, and Water Pollution Control Federation. Standard Methods for the Examination of Water and Wastewater, 14th ed. Washington, DC: American Public Health Association, 1975.

American Water Works Association. A Survey of Operating Data for Water Works in 1960. New York: American Water Works Association, 1964.

Andrews, Wade H.; Madsen, Gary E.; and Hardin, Clay W. Testing Social Indicators in the Techcom Model for Water Development. Logan: Utah State University, 1979.

Atkinson, John W. "British Experiences on Problems of Drinking Water Quality" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

Batchelor, R. A. "Household Technology and the Domestic Demand for Water." Land Economics 3 (August 1975): pp. 208-223.

Baumann, Duane D. and Dworkin, Daniel M. "Consumer Perception and the Choice of Drinking Water Quality" in Safe Drinking Water: Current and Future Problems. Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

Board of Directors, American Water Works Association. Quality Goals for Potable Water. January 28, 1968.

Board of Governors of the Federal Reserve System, Banking and Monetary Statistics 1941-1970, Washington, DC, n.d.

Brown, Robert M.; McClelland, Nina I.; Deininger, Rolf A.; and Tozer, Ronald G. "A Water Quality Index--Do We Dare?" Water and Sewage Works 117 (October 1970):339-343.

Bruvold, William H. "Consumer Attitudes Toward Taste and Odor in Water." Journal of the American Water Works Association (October 1977):562-564.

Bruvold, William H. "Human Perception and Evaluation of Water Quality," Critical Reviews in Environmental Control, Vol. 5, 1975, pp. 153-231.

Bureau of Census. 1970 Census of Housing: Detailed Housing Characteristics, Final Report HC(1)-B1. Washington, DC: US Government Printing Office, July 1972.

Bureau of Census, Statistical Abstract of the US, Washington, DC, 1978.

Bureau of Economic Analysis, The National Income and Product Accounts 1929-1974: Statistical Tables, Washington, DC: GPO, 1974.

Bureau of Economic Analysis, Survey of Current Business, Washington, DC, July 1979.

Bureau of Water Hygiene, US Environmental Protection Agency, Region IV. Evaluation of the Kentucky Water Supply Program: Summary. US Environmental Protection Agency, May 1972.

Bureau of Water Hygiene, US Environmental Protection Agency, Region IV. Evaluation of the Tennessee Water Supply Program. US Environmental Protection Agency, January 1971.

Cabelli, Victor J., and McCabe, Leland J. "Recreational Water Quality Criteria," *News of Environmental Research in Cincinnati.* Cincinnati: National Environmental Research Center, November 11, 1974.

Calabrese, Edward J. *Methodological Approaches to Deriving Environmental and Occupational Health Standards*. New York: John Wiley & Sons, 1978.

Calabrese, Edward J.; Tuthill, Robert W.; Sieger, Thomas L.; and Klar, Janelle M. "Lead and Cadmium Contamination During Acid Preservation of Water Samples." *Bulletin of Environmental Contamination and Toxicology* 23: No.1/2 (September 1979).

Campbell, Michael D., and Lehr, Jay H. *Water Well Technology.* New York: McGraw-Hill Book Company, 1973.

Clark, Robert M. "Water Supply Regionalization: A Critical Evaluation." *Journal of the Water Resources Planning and Management Division,* American Society of Civil Engineers, Vol. 105: WR 2 (September 1979): 279-294.

Danielson, Leon E. *Estimation of Residential Water Demand.* Raleigh: North Carolina State University, 1977.

Dee, Norbert; Baker, Janet K.; Drobny, Neil L.; Duke, Kenneth M.; and Fahringer, David C. "Environmental Evaluation System for Water Resource Planning." Report No. PB 208 822 Battelle Columbus Laboratories, Columbus, Ohio (January 1972).

Deininger, Rolf A. and Landwehr, Jarate M. "A Water Quality Index for Public Water Supplies." Unpublished report, Department of Environmental and Industrial Health, School of Public Health, University of Michigan, Ann Arbor, Michigan (July 1971).

Dinius, S. H. "Social Accounting System for Evaluating Water Resources." *Water Resources Research* 8(5)(October 1972):1159-1177.

Doerksen, Harvey R. and Pierce, John C., editors. "Citizen Influence in Water Policy Decisions: Context, Constraints and Alternatives" in *Water Politics and Public Involvement.* Ann Arbor: Ann Arbor Science Publishers, Inc., 1976.

Drexel University and National Science Foundation. *Water Quality and Health Significance of Bacterial Indicators of Pollution—Workshop Proceedings.* Wesley O. Pipes, ed. Philadelphia: Drexel University, 1978.

XIII-42

Engineering News-Record. A weekly newspaper. New York: McGraw-Hill Publishing Company.

Federal Register, Vol. 45, No. 168, August 27, 1980, p. 57336.

Gallup Organization, Inc. *Public Attitudes Toward the Quality of Drinking Water.* Princeton, NJ, 1974. Also, National Survey Conducted for the American Water Works Association. Princeton, NJ, 1973.

Geldreich, Edwin E. "Applying Bacteriological Parameters to Recreational Water Quality." *Journal of the American Water Works Association* (February 1970): 113-120.

Geldreich, Edwin E. "Fecal Coliform and Fecal Streptococcus Density Relationships in Waste Discharges and Receiving Waters." *CRC Critical Reviews in Environmental Control* (October 1976):349-368.

Gibbs, Kenneth C. "Price Variable in Residential Water Demand Model." *Water Resources Research* 14:1 (February 1978): pp. 15-18.

Glaze, William H.; Saleh, Farida Y.; and Kinstley, Warren. "Characterization of Nonvolatile Halogenated Compounds Formed During Water Chlorination." In *Water Chlorination: Environmental Impact and Health Effects,* Vol. 3, edited by Robert L. Jolley, William A. Brungs, Robert B. Cummings, and Vivian A. Jacobs, pp. 99-108. Ann Arbor: Ann Arbor Science Publishers, Inc., 1980.

Grant, Neville. "Mercury in Man." *Environment* 13:4(May 1971):2-15.

Greely, R. S.; Johnson, A. J.; Rowe, W. D.; and Truett, J. B. "Water Quality Indices." Report No. M72-54, MITRE Corporation, McLean, Virginia (April 1972).

Grima, A. P. *Residential Demand, Alternative Choices for Management.* Toronto: University of Toronto Press, 1972.

Grunewald, O. C.; Haan, C. T.; Debertin, David L.; and Carey, D. I.; "Rural Residential Water Demand: An Econometric and Simulation Analysis." *Water Resources Bulletin* 12:5 (October 1976): pp. 951-961.

Hanke, Steve H. "Demand for Water Under Dynamic Conditions." *Water Resources Research* 6:5 (October 1970): pp. 1253-1261.

Harkins, Ralph D. "An Objective Water Quality Index." *Journal of the Water Pollution Control Federation* 46 (March 1974): 588-591.

Hem, John D. *Study and Interpretation of the Chemical Characteristics of Natural Water*. Washington, DC: US Government Printing Office, 1970.

Horton, Robert K. "An Index-Number System for Rating Water Quality." *Journal, Water Pollution Control Federation* 37 (March 1965):300-306.

Howe, Charles W., and Linaweaver, F. P. "The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structure." *Water Resources Research* 3:1 (First Quarter 1967): pp.13-32.

Inhaber, H. "Environmental Quality: Outline for a National Index for Canada." *Science* 186 (November 1974):798-805.

Johanson, Edward E. and Johnson, Jaret C. "Identifying and Prioritizing Locations for the Removal of In-Place Pollutants." Prepared for the US Environmental Protection Agency, Washington, DC, Contract No. 68-01-2920 (May 1976).

Jonasson, I. R., and Boyle, R. W. "Geochemistry of Mercury." Special Symposium on Mercury in Man's Environment, Ottawa, Canada, February 1971.

Levin, Morris A. "Significance of Wastewater Disinfection to Health Effects Observed in Swimmers." In *Water Chlorination: Environmental Impact and Health Effects*, pp. 11-25.

McCabe,Leland J.; Symons, James M.; Lee, Roger D.; and Robeck, Gordon G. "Survey of Community Water Supply Systems." *Journal of the American Water Works Association* (November 1970):670-687.

McCaull, Julian. "Building a Shorter Life." *Environment* 13:7 (September 1971): p. 3 ff.

McCaull, Julian, and Crossland, Janice. *Water Pollution*. New York: Harcourt Brace Jovanovich, 1974.

McDuffie, Bruce and Haney, Jonathon T. "A Proposed River Pollution Index." Paper presented at the spring 1973 meeting of the American Chemical Society, Division of Water, Air, and Waste Chemistry, New York (April 1973).

McKee, Jack Edward, and Wolf, Harold W., eds. *Water Quality Criteria*, 2nd ed. Sacramento: State of California State Water Quality Control Board, 1963.

McKetta, John J., and Cunningham, William A. *Encyclopedia of Chemical Processing and Design*, Vol. 5, pp. 406-421. New York: Marcel Dekker, Inc., 1977.

Midwest Planning Service. *Private Water Systems*. Ames, Iowa: Iowa State University, 1968.

National Research Council. *Drinking Water and Health.* Washington, DC: National Academy of Sciences, 1977.

National Research Council, Subcommittee on Iron of the Committee on Medical and Biological Effects of Environmental Pollutants, Division of Medical Sciences Assembly of Life Sciences. *Iron*, p. 112. Baltimore, MD: University Park Press, 1979.

National Technical Advisory Committee to the Secretary of the Interior. *Water Quality Criteria*. Washington, DC: US Government Printing Office, 1968.

Nemerow, Nelson L. and Sumitomo, Hisashi. "Benefits of Water Quality Enhancement." Syracuse University, Syracuse, New York, Report No. 16110 DAJ, prepared for the US Environmental Protection Agency (December 1970).

O'Connor, Michael F. "The Application of Multi-Attribute Scaling Procedures to the Development of Indices of Water Quality." Ph.D. Dissertation, University of Michigan, University Microfilms No. 72-29, 161 (1972).

Ott, Wayne R. *Environmental Indices.* Ann Arbor: Ann Arbor Science, 1978.

Patterson, Clair C., and Settle, Dorothy M. "Lead in Albacore: Guide to Lead Pollution in Americans." *Science* 207 (March 14, 1980):1167-1176.

Pipes, Wesley O. (editor). *Water Quality and Health Significance of Bacterial Indicators of Pollution.* Proceedings of a workshop held at Drexel University, Philadelphia, Pennsylvania. April 17-18, 1978.

Prasad, Ananda S. *Trace Elements and Iron in Human Metabolism*. New York & London: Plenum Medical Book Co., 1978.

Prati, L.; Pavanello, R.; and Pesarin, F. "Assessment of Surface Water Quality by a Single Index of Pollution." *Water Research* 5 (1971):741-751.

Revolution in U.S. Farming? Washington, DC: US Department of Agriculture, 1979.

Rohlich, Gerard A. "Drinking Water and Health" in *Safe Drinking Water: Current and Future Problems.* Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

Russell, Clifford S., editor. *Safe Drinking Water: Current and Future Problems, Proceedings of a National Conference in Washington, DC*, 1978.

Safe Drinking Water Act--As Amended through November 1977, printed for the Senate Committee on Environment and Public Works, serial 95-10. Washington, DC: US Government Printing Office, 1977.

Safe Drinking Water Act (Serial No. 95-10). Washington, DC: US Government Printing Office, 1977.

Sale, Kirkpatrick. *Human Scale.* New York: Coward, McCann and Geoghegan, 1980.

Saunders, Robert J. and Warford, Jeremy J. *Village Water Supply: Economics and Policy in the Developing Third World.* Baltimore: Johns Hopkins University Press, 1976.

Schaeffer, David J. and Janardan, Konanur G. "Communicating Environmental Information to the Public: A New Water Quality Index." *Journal of Environmental Education* 8 (Summer 1977):18-26.

Schertz, Lyle P.; McArthur, W. C.; Lagrone, William Franklin; Johnson, Stanley S.; Jesse, Edward V.; and Reimund, Donn A., in Schertz, Lyle P. *et al. Another* Shearer, Lois Ann; Goldsmith, John R.; Young, Clarence; Kearns, Owen A.; and Tamplin, Benjamin R. "Methemoglobin Levels in Infants in an Area with High Nitrate Water Supply." *American Journal of Public Health* 62 (September 1972): 1174-1180.

Seidel, Harris F. and Robert E. Baumann. "A Statistical Analysis of Water Works Data for 1955." *Journal of the American Water Works Association* XLIX (December 1957).

Sontheimer, Heinrich. "European Experience with Problems of Drinking Water Quality" in *Safe Drinking Water: Current and Future Problems.* Clifford S. Russell, editor. Washington, DC: Resources for the Future, 1978.

XIII-96

Spaulding, Irving A. "Social Class and Household Water Consumption." Social Behavior, Natural Resources, and the Environment. New York: Harper & Row, 1972.

Steiss, Alan Walter. Local Government Finance, Lexington, Massachusetts, Lexington Books, 1975.

Stevenson, Albert H. "Studies of Bathing Water Quality and Health." American Journal of Public Health 43:529-538 (May 1953).

Stoner, Jerry D. "Water Quality Indices for Specific Water Uses." US Geological Survey, Reston, Virginia, Circular No. 770 (1978).

Temple, Barker & Sloane, Inc. Survey of Operating and Financial Characteristics of Community Water Systems. Submitted to US Environmental Protection Agency, Washington, DC. Wellesley Hills: Temple, Barker & Sloane, Inc., 1977.

Thomas, W. I. The Polish Peasant in Europe and America. New York: Dover Press, 1958.

Truett, J. B.; Johnson, A.C.; Rowe, W. D.; Feigner, K. D.; and Manning, L. J. "Development of Water Quality Management Indices." Water Resources Bulletin 11(3)(June 1975):436-448.

US Department of Agriculture. Economics, Statistics and Cooperative Service. Farm Income Statistics, Statistical Bulletin No. 627, September 1979.

US Environmental Protection Agency. Bottled Water Study: A Pilot Study of Water Bottlers and Bottled Water. Washington, DC: Water Supply Division, Environmental Protection Agency, 1972.

US Environmental Protection Agency. "Drinking Water Regulations: Radionuclides." Federal Register 41: No. 133 (July 9, 1976):28402-28405.

US Environmental Protection Agency. "Interim Primary Drinking Water Regulations—Control of Organic Contaminants in Drinking Water." Federal Register 43: No.28 (February 9, 1978):5756-5780.

US Environmental Protection Agency. Manual for Evaluating Public Drinking Water Supplies. Washington, DC: US Environmental Protection Agency, 1975.

US Environmental Protection Agency. Manual of Individual Water Supply Systems. 2nd rev. ed. Raleigh, NC: US Environmental Protection Agency, 1974.

US Environmental Protection Agency. National Interim Primary Drinking Water Regulations (EPA-570/0-76-003). Washington, DC: US Government Printing Office.

US Environmental Protection Agency. "National Interim Primary Drinking Water Regulations." Federal Register 40: No. 248 (December 24, 1975), 59566-59574.

US Environmental Protection Agency, National Safe Drinking Water Strategy: One Step at a Time. Washington, DC, February 1977.

US Environmental Protection Agency. "National Secondary Drinking Water Regulations," Federal Register 42: No. 62 (March 31, 1977), 17144-17146.

US Environmental Protection Agency, Region I. Evaluation of the Connecticut Water Supply Program. US Environmental Protection Agency, September 1974.

US Environmental Protection Agency, Region X. Idaho Water Supply Program Evaluation: Summary. Seattle, Washington: US Environmental Protection Agency, October 1973.

US Environmental Protection Agency. State of the Art of Small Water Treatment Systems. Washington, DC: US Environmental Protection Agency, 1977.

Walski, Thomas M. and Parker, Frank L. "Consumers Water Quality Index." Journal of the Environmental Engineering Division ASCE (June 1974):593-611.

Water Supply Program, US Environmental Protection Agency, Region VII. Evaluation of the Kansas Water Supply Program. US Environmental Protection Agency, undated (study conducted 1972).

Water Supply Section, US Environmental Protection Agency, Region VIII. Evaluation of the Colorado Water Supply Program. Denver: US Environmental Protection Agency, November 1974.

Whitsell, Wilbur J.; Hutchinson, Gary D.; and Taylor, Donald H. "Problems of Individual Supplies." The State of America's Drinking Water. Raleigh: North Carolina State University, 1975.

Whyte, William Foote. Organizing for Agricultural Development. New Brunswick, NJ: Transaction Books, 1975.

Wood, John M. "A Progress Report on Mercury," *Environment* 14:1(January/February 1972): 33-39.

World Health Organization. *European Standards for Drinking-Water*. Geneva: World Health Organization, 1970.

World Health Organization. *International Standards for Drinking-Water*. Geneva: World Health Organization, 1958.

Wright, Forrest B. *Rural Water Supply and Sanitation* (Third Edition). Huntington, New York: Robert E. Krieger Publishing Company, 1977.

Zoeteman, B. C. "The Potential Pollution Index as a Tool for River Quality Management." Technical Paper No. 6, Government Institute for Water Supply, World Health Organization, International Reference Centre for Community Water Supply, The Hague, The Netherlands (August 1973).